



State of North Dakota Business Intelligence Assessment:

Findings and Recommendations

ciber[®]

State & Local

CIBER Inc.

June 1, 2006

Copyright Notice

This document (written or otherwise displayed on magnetic media or other medium) contains confidential and proprietary information of the state of North Dakota.

It has been furnished only for informational purposes, and no license or permission is hereby granted to use such information in any manner.

In no event may this information be reproduced, distributed and/or publicly displayed in any form or by any means without prior expressed written permission of the state of North Dakota. All rights reserved.

Table of Contents

Executive Summary	5
Key Findings	6
Recommendations	6
Next Steps	9
Conclusion	10
Business Intelligence Overview	13
Business Intelligence Overview	13
Business Impact Opportunities	14
State Government Examples of BI Applications	15
Data Mining and Advanced Analytics Overview	18
Text Mining Overview	19
Data Integration/ETL (extract, transform, load) Overview	20
Assessment of Current North Dakota Business Intelligence	21
Inventory of State Business Intelligence Projects	21
Current Business Intelligence Related State Technology Standards	23
Assessment of State Business Intelligence Practices	23
BI Assessment Approach	24
BI Analysis Approach	24
Key Resources	24
Key Findings	25
Recommendations	26
1. Enterprise Business Intelligence Software Standard	26
Rationale for Selecting Cognos and SAS	26
Comparative Ratings: SAS vs. Cognos	27
The Other Business Intelligence Software vendors	28
Using Cognos as the Primary BI Solution	28
Using SAS for Data Mining and Advanced Analytics	30
Sharing MetaData Between SAS and Cognos	31
2. Enterprise Data Integration/ETL Software	32
State government data integration/ETL case study	32
The Challenge	32
The Solution	33
The Results	34
3. Enterprise Data Warehouse Platform Standard	35
4. Integrate BI with the State's Enterprise Architecture	37
5. Replace Non-Standard BI Software	42
6. Perform BI Assessments of New Application Systems	42
7. Establish a BI Competency Center	42
8. Consider using Cognos BI for the Peoplesoft BI applications	45
9. Consider BI Cost Sharing with Other Organizations	46
10. Establish a Statewide BI Application Advisory Council	48
11. Utilize Consultants	49
12. Establish a Budget	49

Appendix A: Current North Dakota BI Projects52
Appendix B: Business Intelligence Solution Requirements54
Appendix C: North Dakota BI Enterprise Architecture Chart (a sample view of potential enterprise application integration)56
Appendix D: Glossary.....57

Executive Summary

North Dakota's current economic development slogan is "*North Dakota: a new state of business.*" This slogan reflects North Dakota's desire to build a better future for its citizens and businesses by strategically positioning the state for future prosperity in the face of a competitive global marketplace.

This mission requires that North Dakota leaders make wise decisions regarding future strategies and allocation of state funds. Wise decisions require timely, accurate business intelligence. Timely, accurate business intelligence requires effective business intelligence technology and proven business intelligence best practices. This report provides recommendations for building an effective and reliable business intelligence capability for the state of North Dakota. To borrow from the State's economic development slogan, these recommendations represent a "*new state of business intelligence*" for North Dakota.

In developing the business intelligence recommendations in this report, two key questions were addressed:

- How can the state of North Dakota best deploy the use of business intelligence (BI) technology?
- Which business intelligence tools can best serve the State's unique business intelligence needs and internal culture?

Background and Objectives

In April, 2006, the state of North Dakota's Business Intelligence Study team selected CIBER, Inc. to conduct a high level (118 hour) assessment of the State's business intelligence architecture and strategy. The primary deliverables for this project were to provide the State with:

- A compilation of the State's business intelligence requirements
- Specific vendor/technology pros and cons related to the business intelligence requirements
- Recommended business intelligence and data warehousing information architecture
- High-level business intelligence and data warehousing implementation strategy
- High level implementation budget

Approach

To fulfill the Business Intelligence assessment deliverables, CIBER utilized the following approach:

- Reviewed research already gathered by the State's Business Intelligence Study team.
- Analyzed the State's unique business needs through on-site interviews with key State employees, and an assessment of key State documents (Budget document, IT Strategic Plan, agency website content).
- Mapped the assessment findings against CIBER's extensive business intelligence knowledge base and experiences (CIBER is recognized by Gartner Group as one of the top 20 business intelligence consultancies in the world), and against CIBER's extensive state government experience (CIBER has worked with 46 of the 50 U.S. states).

- Developed a report of preliminary findings and recommendations that was presented and reviewed on-site with key State representatives.
- Prepared a final report of findings and recommendations for utilizing business intelligence at the state of North Dakota (represented by this document)

Key Findings

Based on the assessment of the State’s business intelligence architecture, processes, and directions, the following key findings were identified:

- **Multiple BI Software Products in Use.** Currently, the State is using several different BI related software products. Vendors represented by these various BI software products include SAS, Oracle, and Microsoft
- **Multiple Data Integration (ETL) Software Products in Use or Planned.** The State owns or is considering the use of multiple data integration/ETL (extract, transform, load) software products. Vendors represented by these software products include IBM (formerly Ascential), Informatica, and Data Mirror.
- **No statewide BI coordination or strategy.** There are no statewide strategies or plans for coordinating the development or management of BI applications in the State. This has resulted in several “siloes” pockets of BI applications, and the use of multiple BI tools.
- **Limited in-house BI expertise and skills.** The State has limited internal BI expertise due primarily to the fact that outside consultants performed most of the work with current BI applications which included only limited internal BI training and knowledge transfer.
- **No formal BI methodologies.** There are no formal BI methodologies being used by the State in the development or operation of BI applications.
- **No formal BI requirements for new application development projects.** There are no BI requirements linked to the development of new application systems or to the purchase of new application software.
- **The existence of a large amount of legacy data (including Adabas and DB2 data).** The state has a large number of legacy applications that have data residing in older database formats (e.g. Adabas). The State’s current BI tools are not adequate for accessing and using legacy data for future business intelligence applications.
- **Business intelligence is underutilized.** The State has significant opportunities to increase the use of business intelligence technology to support critical State decision-making and operations.

Recommendations

Based on the key findings presented above, CIBER has developed twelve specific recommendations for the State:

1. **Adopt an enterprise Business Intelligence software standard.** CIBER recommends that the State adopt the Cognos Business Intelligence Suite as the primary enterprise-wide business intelligence software for statewide use. In addition, CIBER recommends that the State consider adopting the SAS Business Intelligence Suite as a complementary statewide standard for

deployment in special situations where advanced data mining, predictive analytics, text mining, and geographical (spatial) integration with the ESRI GIS system are key considerations.

2. **Adopt an enterprise Data Integration (Extract, Transform, Load – ETL) software standard.** CIBER recommends that the State select a standard data integration software suite for statewide use. CIBER recommends that the State consider the use of Informatica or SAS as the state standard. Both SAS and Informatica offer top rated data integration/ETL products that can integrate well with the State’s legacy databases (e.g. Adabas). In addition, SAS is also recommended as the State’s secondary BI software standard, and Informatica may be acquired anyway if the State legislature approves the new Medicaid Management Information System (MMIS) option that includes the use of the Informatica software. Since the MMIS decision is not expected until spring 2007, it is recommended that the State wait until that time before making a final decision on the selection of a State standard data integration/ETL software solution. Until that time, it is projected that the next BI application (PeopleSoft) could use the Cognos ETL functionality for the existing PeopleSoft ETL functionality (see Recommendation #8 below).
3. **Adopt an enterprise Data Warehouse Platform standard.** CIBER recommends that the State consider the purchase and use of the Netezza data appliance as the primary statewide data warehouse platform. By architecturally integrating database, server, and storage within a single appliance, the Netezza data appliance can deliver 10 to 100 times the performance at half the cost of traditional data warehouse platforms using general purpose database software. In those cases where a data appliance may not be practical, the use of the Oracle 10G database is recommended as the data warehouse repository standard for the state.
4. **Integrate the recommended BI solution into the State’s Enterprise Architecture.** CIBER recommends that the State’s BI standards be integrated as part of the State’s overall Enterprise Architecture strategy for data collection, sharing, and reporting. The proposed standard BI and data integration/ETL software should be considered as an integral part of the total enterprise computing infrastructure for the State. This would include the potential integration of the BI solution with the State’s FileNet document management system, the State’s Microsoft Active Directory standard, and the State’s ESRI GIS system. (A graphic description of how the proposed State BI Architecture might potentially fit into the State’s Enterprise Architecture is shown in Appendix C).
5. **Eventually replace the current non state standard BI related software with state standard BI software.** CIBER recommends that the State eventually replace current non standard BI related software with the new proposed State standard BI and data integration/ETL software. This would include removing the Business Objects Crystal Reports software and Oracle Discover where and when it is practical to do so. Removing non State standard BI and ETL software will reduce software and support costs, improve the quality of data sharing, and help promote “one version of the truth.
6. **Perform BI assessments of new application systems.** CIBER recommends that the State perform BI impact assessments for all new application systems developed or purchased by the State. These assessments would identify potential benefits for integrating the proposed new application systems with the State’s BI and ETL software standards.

7. **Establish a central Business Intelligence Competency Center.** CIBER recommends that the State establish a central BI Competency Center to minimize costs and maximize quality in the deployment and support of BI applications. The BI Competency Center would house the administrators for the State standard BI software (Cognos and SAS) and the State standard data integration/ETL software (Informatica or SAS). The Competency Center would work with vendors, outside consultants, and state employees as needed to provide coordination and promote best practices. An associated statewide BI User Group is also recommended along with a BI User web site for posting useful BI tips, FAQs, and lessons learned.
8. **Consider using the Cognos BI solution for the PeopleSoft business intelligence applications.** While the State owns the PeopleSoft EPM software, it is recommended that the State consider using the Cognos BI tool for the PeopleSoft BI applications. Use of the Cognos BI solution will make it easier to link and integrate PeopleSoft data with non PeopleSoft application data at the State.
9. **Consider sharing BI costs with other organizations.** While the focus of this BI assessment was on the State of North Dakota, it is recognized that other State related organizations could also potentially benefit from sharing the proposed State BI resources (hardware, software, and competency center). Consequently, it is recommended that the State consider sharing some of the costs of the BI Competency Center, the Data Warehouse platform (Netezza), and the BI and data integration/ETL software with other organizations – such as the Bank of North Dakota, the North Dakota University system, and local governments and school districts. In addition, some of the BI costs may be eligible for reimbursement with Federal funds in cases where State BI resources are used to support federally funded programs.
10. **Establish a statewide BI Application Advisory Council.** The State should consider the establishment of a Statewide BI Application Advisory council of senior State officials to identify and prioritize the development of business intelligence applications for the State. This council could also be a subset of existing State IT advisory committees or groups. It is recommended that this council include elected officials and the State’s Chief Information Officer.
11. **Utilize outside consultants on an as needed basis.** The State should utilize outside consultants from the BI and data integration/ETL software vendors and/or other BI consultancies as a means of reducing risks, speeding deployment, transferring knowledge, and reducing costs. Because of the State’s limited BI and data integration/ETL expertise, use of outside consultants are recommended to facilitate BI knowledge transfer and to help establish the BI Competency Center.
12. **Establish a budget.** Based on the recommendations in this report, CIBER recommends that the State establish a budget of approximately \$1,240,000 to cover the recommendations offered in this report. This recommended budget is based on the following cost estimates:
 - **BI Software– \$500,000.** Includes Enterprise licenses, first year support costs, administrator training, and software installation for both Cognos and SAS. This cost also includes the use of the MITI Meta Integration® Model Bridge (MIMB) to integrate Meta data between SAS and Cognos on an as needed basis.

- **Data Integration/ETL Software - \$240,000.** This would cover the one time purchase cost, first year support costs, administrator training, and installation costs for the purchase of data integration/ETL software (Informatica or SAS).
- **Data Warehouse Platform - \$270,000.** This would include the purchase of a data appliance (Netezza) that would include hardware and data storage with backup to support two terabytes of data warehouse data. This cost includes the one time purchase cost, first year maintenance fee (18%), installation, and onsite administrator training.
- **Training and Consulting - \$150,000.** This would cover the costs of using outside consultants as needed to establish the BI Competency Center and to assist with issues and problems associated with building initial BI and ETL applications. This would also cover travel and registration costs for key BI Competency Center staff to attend two key BI conferences and two BI/ETL vendor user conferences.
- **Server hardware - \$80,000.** This would cover the costs of two processor servers and related hardware to house the BI Software and the data integration/ETL software.
- **State employee costs (reassignment of five existing staff).** It is recommended that five individuals be initially assigned to work part time in the BI Competency Center: a BI software administrator, a Data Integration administrator, a business intelligence analyst, a database administrator, and a BI Competency Center director.

NOTE: This budget does not cover the cost of developing individual agency business intelligence applications. Also, this budget does not allocate any funds for the existing state personnel that would be assigned part time or full time to work in the BI Competency Center.

Next Steps

In implementing the BI recommendations, the following next steps are offered for consideration by the State of North Dakota:

- **Form the proposed BI Application Advisory Council** to prioritize business intelligence applications for the State for the next three years, and to secure appropriate funding as part of the budget process.
- **Conduct a proof of concept test with the Cognos software for the PeopleSoft BI application.** Since the next State BI application is projected to be the PeopleSoft BI application, it is recommended that the State conduct a proof of concept with the selected enterprise BI software vendor (Cognos) to validate the appropriateness of using the Cognos software for the PeopleSoft BI applications.
- **Conduct a proof of concept test with the Netezza data appliance.** Also in conjunction with the PeopleSoft BI application, conduct a proof of concept test between the Netezza data appliance and the use of the Oracle 10G data base server for use as the preferred State standard data warehouse platform.
- **Select and engage a BI consultant firm(s)** as needed to assist in the following efforts:
 - The design and deployment of the BI Competency Center

- The procurement of recommended BI and data integration hardware and software (including negotiating with the vendors to obtain maximum benefits in pricing and usage conditions).
- The development of an appropriate BI methodology for managing the BI infrastructure, processes, policies, and applications.
- **Select a State standard data integration/ETL software package.** In conjunction with the results of the MMIS decision expected in the spring of 2007, conduct an evaluation of the recommended data integration software vendors (SAS and Informatica) and make a final selection of a State standard data integration/ETL software tool. Prior to the selection of a primary data integration/ETL software standard in March, the State could use the Cognos or PeopleSoft data integration/ETL capabilities if needed.

Conclusion

The potential Return on Investment (ROI) for the proposed business intelligence strategy in North Dakota is significant. Nearly every State agency will potentially benefit from the use of business intelligence applications.

The high potential ROI value of BI technology is supported by a recent national survey of Chief Information Officers (CIO's) by the Gartner Group. In this 2006 released survey, the deployment of business intelligence technology was ranked as the number one information technology priority in America today.¹

To maximize the ROI value of BI applications in the State, CIBER recommends that the State adopt statewide enterprise standards for BI software, data integration/ETL software, and the data warehouse platform. In addition, CIBER recommends that the State establish a Business Intelligence Application Advisory Council, a statewide BI Competency Center, and a BI Deployment Methodology that incorporate industry best practices.

These recommendations are designed to promote the following benefits for the State of North Dakota:

BETTER deployment of future BI applications

- Improved prioritization of BI applications. Use of the proposed BI Advisory Council will help ensure that State BI projects are prioritized on the basis of strategic importance and total ROI to the State.
- Improved access to State data sources. Use of the proposed data integration/ETL software will make it easier and less costly to access a wide variety of State data sources – particularly legacy data systems.
- Integration with FileNet and ESRI GIS applications. Use of the proposed data integration/ETL and BI software will make it easier to integrate BI applications with the State's FileNet document management applications, and the ESRI GIS mapping applications.
- Improved BI usage and satisfaction. Use of BI Competency Centers such as the one recommended for the State, have been shown to improve BI usage by 74%, BI user satisfaction by 48%, and increased decision-making speed by 45%.²

- Improved data quality. Use of a central shared BI metadata (data about data) can promote shared understanding of reported data and help promote a single “version of the truth.”
- Improved master data management. Use of the proposed data integration/ETL software will make it easier and less costly to deploy master data management applications.

FASTER deployment of future BI applications

- Faster data integration. Use of data integration software can reduce the time to build data integration applications (extract, transform, and load data from source systems to the data warehouse) over manual methods.
- Faster development of BI applications. Use of shared data and automated BI tools can reduce the time required to build BI applications over manual methods.

SAFER deployment of future BI applications

- Fewer BI deployment errors. Use of the BI Competency Center can help avoid costly errors by focusing on best practices and using a core group of trained BI professionals.
- Fewer BI and data integration/ETL software errors. Use of experienced consultants can help reduce BI and data integration/ETL software errors.
- Fewer data quality problems. Use of a high quality data integration/ETL tool can do a better job of profiling source data (particularly legacy data systems) and cleaning “dirty” data than manual methods without data integration/ETL software.

LESS COSTLY deployment of future BI applications

- Less costly BI application development costs. The use of data integration software may reduce the time to build BI applications on average by 25 to 50% over manually coded data integration/ETL solutions.
- Reduced IT infrastructure costs. Use of the recommended data warehouse platform will reduce costs of deploying BI applications. The proposed Netezza solution claims that its data appliance offers a minimum of 10 times improved processing performance for 50% less than the cost of general purpose data base platforms – such as Oracle’s DBMS or IBM’s DB2.
- Reduced BI support and software costs. Use of a BI Competency Center can potentially reduce BI staff costs by 26% and reduce BI software costs by 24%.³

In addition to the benefits above, the primary benefit of the proposed BI solution is to promote better decision-making. Opportunities for improved decision making include: improving student education, improving healthcare, attracting new jobs, improving operational efficiencies, promoting the success of state industries, and improving public safety.

Using BI technology to make better strategic decisions will help North Dakota compete with other states in the global marketplace. The recent book from Peter Fingar, *Extreme Competition – Innovation and the Great 21st Century Business Reformation*, is a warning call to those states that fail to innovate in a global economy. Business intelligence applications are critical for finding new innovative ways to address state government challenges, and to help promote a prosperous future for North Dakota’s citizens and businesses.

“The key to winning in the Information Age is making decisions that are consistently better and faster than the competition – survival of the fittest. Business intelligence is an approach to

managing your business that is dedicated to providing competitive advantage through the execution of fact-based decision making. “

Elizabeth Vitt, Michael Luckevich, and Stacia Misner; *Business Intelligence: Making Better Decisions Faster*, 2002.

“Winning companies will deploy innovative information technology tools to manage human-driven processes, capture information deeply personal to each participant, and help people to use this information both individually and collaboratively. “

Peter Fingar; *Extreme Competition*, 2006.

Business Intelligence Overview

To promote a better understanding of the recommendations made in this report, this section provides a basic overview of key business intelligence and data integration concepts. A glossary of key business intelligence terms is included in Appendix D.

Business Intelligence Overview

If a business intelligence solution can't help you make sound decisions about your company's future – quickly, easily and with confidence – it's neither good business nor intelligent.

The ultimate goal of effective of business intelligence is to promote better decisions faster. This is facilitated by providing software tools and best practices to collect, validate, analyze, and report timely information to key decision-makers – whether the decision maker is a front line employee or the Governor.

A comprehensive Enterprise Business Intelligence solution provides a foundation for addressing these five key questions:

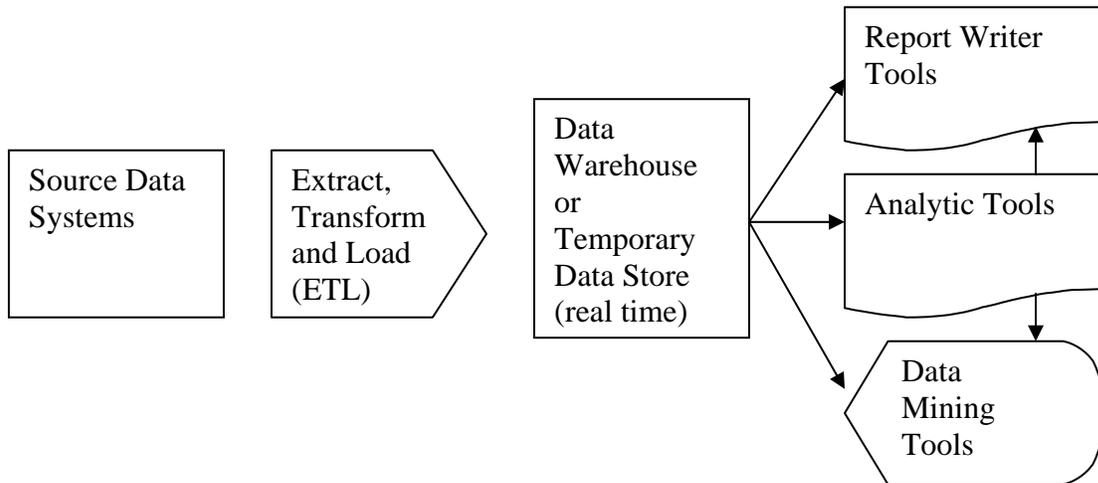
- WHAT happened? (reporting)
- WHY did it happen? (analyzing)
- WHY will it happen? (predicting)
- WHAT is happening? (monitoring)
- MAKING it happen. (event-driven decision support)

Business Intelligence solution technology consists of three primary components:

- **Data Integration technology** (extract, transform, load – ETL) – represents software that is used to analyze source data, clean up any data quality problems, and transfer the data in real time or batch to a data warehouse or temporary data location to be accessed and analyzed by State decision makers
- **Data Warehouse Repositories** (strategic and operational) – represents the computer hardware where extracted data is stored – permanently or in a temporary state for operational analysis or viewing in real time.
- **Business Intelligence Tools** – represents software tools that are used to model and analyze data to support better decisions. Specific business intelligence tools include: basic reporting tools, dimensional analysis tools (OLAP), data mining, dashboards, scorecards, and text mining tools. Types of analysis include: trending, forecasting, “what if” projections, fraud detection, Geographical and/or spatial analysis, and exception reporting.

The relationship of these three BI components can be seen in the exhibit on the following page:

Business Intelligence Infrastructure



The functions and roles typically associated with the development of business intelligence applications include the following:

- **Target decisions.** Identify the key business decisions for which business intelligence is needed. Primary participants: State decision makers – individuals and/or decision teams (e.g. councils or committees).
- **Target data sources.** Identify data sources that can provide the source data needed. Primary participants: source data experts – those most knowledgeable about the source data (can include programmers, system analysts, end users, or database administrators).
- **Extract, transform, and load source data.** Analyze the source data to determine requirements for extracting, transforming (e.g., cleaning dirty data, summarizing detail data, etc.), and transferring data to a target database for subsequent analyst in real time or batch. Primary participants: Data integration (ETL) tool experts – those most familiar with using the data integration software or ETL tools.
- **Decision support modeling.** Develop the logic, models, and display formats by which the source data can be analyzed, mined, correlated, mapped, displayed, and reported. Primary participants – subject matter experts, systems analysts, database administrators, and statisticians.
- **Analysis and reporting.** Develop the actual reports, queries, graphs, dashboards, or scorecards to be used to analyze and display or report the data in the data warehouse. Primary participants – trained users of the BI software tools to build queries and reports – can be business analysts or power users (typically does not require programmers).

Business Impact Opportunities

The strategic use of business intelligence technology offers significant opportunities to promote State goals and improve state government productivity. Specific opportunities include:

1. Off loading management reporting and data analysis from production transactional systems to avoid reducing the performance and response time for the state's production application systems – particularly the State's PeopleSoft financial and human resource systems.
2. Making it easier to find and identify relevant information (both internal and external) that can be quickly and easily accessed by anyone who needs it. This includes promoting “one version of the truth.”
3. Creating an environment where data from multiple system silos can be brought together and analyzed. This is particularly needed to support cross agency data sharing.
4. Providing new capabilities for advanced data analysis – such as forecasting, trending, fraud detection, longitudinal studies, and correlation analysis that is not readily doable with existing State business intelligence technology.
5. Providing critical information that can be used in economic development. This would include doing market analysis for attracting and promoting industry growth in the five target industries that the State Department of Commerce has targeted to broaden the state's economic base: value-added agriculture, advanced manufacturing, information technology, tourism, and energy.

State Government Examples of BI Applications

This section provides actual examples of how some state governments are using business intelligence applications to address key state government challenges.

State of Florida Department of Corrections

As an example of how business intelligence can have a significant business impact in a state government setting, this example from a SAS BI application at the State of Florida is offered (from SAS public website):

A vital component of the state's criminal justice system, the Florida Department of Corrections strives to protect the public by operating a safe, secure, humane, and efficient corrections system. To do this effectively, the department must provide quick and thorough responses to the many information requests it receives from prison and probation officials, prosecutors, state legislators, local police departments, and other government agencies.

That's why it maintains a SAS data warehouse that stores information on every inmate and offender within the state – currently numbering more than 82,000 in custody and more than 150,000 on various types of community supervision. The data warehouse also contains death and employment records, county population demographics, and juvenile justice information. To supplement its own data, the department uses SAS to access, combine and analyze data from other state agencies such as the Department of Law Enforcement, the Department of Juvenile Justice and the Department of Education. Data on the Corrections Department's work force is also maintained, with SAS drawing information from an Oracle database of all state agency staff.

Florida law enforcement agencies rely on the Department of Corrections for ad hoc assistance in narrowing suspect lists for serial crimes. Using SAS to access historical corrections data and analyze it for known criminal patterns and traits, the department can compile reports in a matter of

hours and export them to local police and sheriff departments to assist in the search and apprehension of suspects.

SAS also makes it easy to track activities involving individual inmates, such as transfers, disciplinary actions, and positive drug tests. This information can be used to match random drug testing results with demographic data, for example, so corrections officials can determine the characteristics of drug users within the system and implement new policies to reduce drug use in a prison facility.

Perhaps most important to the citizens of Florida, Ensley's group counts on SAS to create monthly and quarterly reports that track key performance indicators, including an escape report that analyzes details about attempted escapes at each of Florida's 58 correctional institutions. The information has been used to develop departmental standards, to provide information to the state's governor and legislators, and to improve agency operations.

SAS has also been used to predict individual rates of failure for offenders on community supervision, forecasting which individuals are most likely to violate the conditions of supervision set by the court. Program administrators use this information to determine which prisoners will benefit from more intensive supervision.

"The goal of that system is to supervise people on a level commensurate with their risk," explains Ensley. "For the community supervision program, we rank everyone according to their likelihood of failure, and those who are ranked more serious are assigned to officers with lower caseloads, so they can make more personal contact with the offender."

With help from SAS, the Florida Department of Corrections has witnessed the importance of combining, analyzing and interpreting data from throughout the criminal justice system – benefits that ultimately range from detecting and preventing crime to optimizing state resources. "SAS really gives us a scientific basis for allocating our resources," says Ensley. "It allows us to utilize our resources responsibly and in the best way that we can."

New York City Department of Education

An example of how business intelligence can be used for education analysis can be found in New York City's use of Cognos BI tools (information excerpted from a Cognos press release in 2003):

Using Cognos, the NYC Department of Education can look at data collected from across 1,300 schools. By aggregating data from finance, tax levy systems, personnel data, attendance and exam data, the department can determine correlations between class size and student attendance, or between funds allocated for textbooks and student performance. This information is used by administrators across the region to monitor trends in data and scorecard the performance of the students, the schools, and the principals that run them. The information is available via extranet to the general public.

"Using Cognos has greatly increased the speed and accessibility of our data - previously, there was a significant cost and time commitment involved in the manual analysis of this data," said James

Spinelli, consultant to the NYC Department of Education. "By proactively looking at trends across grades and schools we are able to pinpoint problems in the early stages and quickly take corrective action. This level of analysis was simply not possible before. Administrators can now access reports like the Period Attendance Report and enrollment statistics instantly and securely over the Web."

"Cognos provides K-12 educators with the visibility and technology foundation needed to improve student and classroom performance. By understanding the profile of successful classroom performance and investing in the areas that will help students the most, K-12 institutions can build an organization and curriculum that effectively delivers on performance obligations. And above all else, is able to quickly identify students that are falling behind, and assure they get the help needed to assure their success," said Terence Atkinson, director of public sector solutions, Cognos. "Cognos delivers the industry's most complete performance management offering that is best suited to meet the complex needs of the public education systems."

The Minnesota TIES Education Consortium

The Minnesota TIES organization offers another example of using Cognos BI for education analysis (information excerpts from a Cognos press release in 2004):

Working with Cognos Professional Services, Cognos and TIES will offer a performance management solution that includes over 150 pre-built reports for addressing the No Child Left Behind Act of 2002 (NCLB). The NCLB act was implemented to hold states and schools accountable for the academic achievement of all students, and to demand continuous performance improvement from students and teachers.

The TIES and Cognos solution will enable superintendents, principals, teachers and other administrators to analyze and report on student data and receive alerts on pre-determined metrics. Authorized users will be able to look at data multidimensionally with Cognos PowerPlay, drawing information from a central data warehouse, and can analyze and create reports on a number of areas including student performance, test results, curriculum management, attendance and absence reporting, and aggregated performance reviews. These abilities will enable educators to easily identify children who are falling behind, and to empower schools to take specific actions without having to wait for a formal report card to indicate where students are struggling academically.

In addition, school administrators will use Cognos ReportNet for operational and ad-hoc reporting on financial, human resources, and payroll data. Quicker and easier access to information, coupled with the ability to analyze across dimensions will enable schools to examine different aspects of organization performance and how it relates back to student success.

"TIES evaluated several other business intelligence vendors, but after close scrutiny we feel strongly that Cognos has the best solution for our market," said Betty Schweizer, TIES co-executive director. "By leveraging TIES' deep expertise in K-12 education with the powerful reporting, analysis and scorecarding capabilities of the Cognos EBI solution, educators can better understand the driving factors behind student performance and take action at the stage where they

can have maximum impact. In addition, we are offering schools a practical solution that meets the information needs of the No Child Left Behind act, legislation that affects all our school districts."

"We are very pleased to be working with TIES on this important initiative and to continue to broaden our reach in the K-12 market," said Terence Atkinson, director of public sector solutions, Cognos. "TIES has the extensive industry expertise needed to address the student performance challenges faced by school districts in the United States and selecting Cognos is a clear indication of the significant competitive advantage our solution offers to K-12 organizations."

About TIES: TIES, an education technology consortium of 37 Minnesota school districts, provides integrated SIS, finance, and human resource systems. Designed by its members, TIES web-based SIS (TSIS) is one of the most complete systems on the market, incorporating census, health, transportation, IEP, parent access, grade book, test and assessment data, and more. TIES is also one of the largest education technology trainers in the country. For more information about TIES, go to www.ties.k12.mn.us.

Data Mining and Advanced Analytics Overview

Data mining and advanced analytics represents an important subset of business intelligence. Data mining/advanced analytics applications offer some of the highest potential paybacks of using business intelligence applications. This section provides an overview of how advanced analytics (such as predictive analytics) and data mining can benefit state government operations.

Data mining and advanced analytics – if used correctly – can provide significant benefits to state governments in designing and implementing improvement programs. Specific areas with the greatest return on investment potential include:

- Disease prevention
 - What actions/programs can we undertake to best minimize the occurrence of major preventable diseases in our state (obesity, diabetes, hearth disease, cancer)?
- Pandemic Response
 - What back up resources are available to fill vacancies suddenly created by pandemic illnesses?
 - What alternative suppliers are available to fill orders?
 - What hospital beds and medical providers will be needed?
 - What strategic stockpiles will be needed for critical medical supplies?
- Crime prevention
 - What actions/programs can we undertake to minimize crime in our state?
- Offender apprehension
 - What actions/programs can we undertake to apprehend criminals in our state?
- Economic Development
 - What actions/programs can we undertake to attract businesses most likely to benefit from the type of resources our state has to offer?
- Public Education (K-12)
 - What actions/programs can we undertake to minimize high school dropouts?
- Offender Reform

- What actions/programs can we undertake to help offenders become productive citizens and not return to jail or prison?
- Revenue Maximization
 - What actions/programs can we undertake to make sure all businesses and citizens are paying taxes due to the state?
- Fraud Detection
 - What actions/programs can we undertake to identify fraudulent activity by providers or recipients in state welfare and healthcare programs
- Procurement Optimization
 - What actions/programs can we undertake to maximize competitive bidding on high dollar products and services purchased by the state?
- Traffic Optimization
 - What actions/programs can we undertake to minimize traffic jams and traffic accidents in our state?

Of the opportunities listed above, the area of healthcare represents one of the highest payback opportunities. Today, healthcare costs represent the fastest and largest growing cost of state governments – primarily in the area of Medicaid. New solutions are needed to control healthcare costs. Data mining and predictive analytics are considered two essential tools for finding new, innovative solutions.

“Never before has disease prevention and health promotion been more important. As world events develop, with the added threat of bio-terrorism and other emerging public health issues, those who can apply knowledge gained through research to real-world problems are in great demand across all sectors: healthcare, pharmaceuticals, governmental & non-governmental agencies, business and academia.”

Drexel University elearning web site 05-22-06.

Text Mining Overview

Text mining represents another subset of business intelligence. Text mining involves analyzing unstructured data – such as data found in word processing documents, emails, text messages, and PDF documents. It is estimated that more than 85 percent of all business information exists as unstructured data -- commonly appearing in e-mails, memos, notes from call centers, news, user groups, chats, reports, letters, surveys, white papers, marketing material, research, presentations, and Web pages.”⁴

The ability to mine unstructured data for relevant information fills a major gap in the deployment of business intelligence to improve decision making. Here is just one example of how text mining combined with traditional business intelligence can be used to improve outcomes:

UAB Health System selected SAS Enterprise Miner™ and SAS Text Miner to analyze information from doctors’ notes and textual comments in its patient database. SAS delivers advanced analytic technology that will convert unstructured data into intelligence that can be mined in concert with demographic variables to track and predict patient response to various medical procedures. The UAB Health System is looking for patterns and clusters of repeated indicators to point to the best methods of treatment.

“We plan to use this study to develop an early-warning system that supports physicians,” Hardin said. “Physicians see patients and provide dictation in electronic format along with clinical encounter notes and other textual information. We seek a way to combine information from medical transcriptions with lab results and other numerical data to create a full picture of patient health and to help our physicians make better clinically based decisions.”

According to Hardin, the study may suggest actions that administrators can take to reduce hospital stays and expenses, improve overall patient care and satisfaction, and lead to more accurate diagnoses. And study organizers hope the findings will extend far beyond the UAB Health System.⁵

Data Integration/ETL (extract, transform, load) Overview

The process of identifying, collecting, cleansing, and transforming source data into a state for business intelligence processing is known as data integration or extract, transform, and load (ETL). The data integration/ETL process typically represents 70% of the total cost of a business intelligence application development project.

Without data integration/ETL software tools, the data integration process can be excessively costly in both time and money. There is also a greater chance that bad data will slip into the data warehouse and data marts.

The consequences of using bad data for decision making are serious. Damage caused by using bad data in decision making can include:

- disgruntled citizens and businesses
- lost revenue
- law suits
- reduced confidence in BI applications
- costs associated with resending erroneous data reports
- political embarrassment
- breaches of privacy
- breaches of regulatory compliance (such as HIPAA)

Because of these potential problems, the use and selection of data integration/ETL software is considered critical. This is not an area where corners should be cut.

In a 2002 survey conducted by The Data Warehousing Institute, almost half (44%) of the respondents said the quality of the data within their companies was “worse than everyone thinks” (*TDWI Report Series, Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data*, by Wayne Eckerson, January 2002). According to TDWI’s Data Quality Survey (mentioned above), almost half of the surveyed companies (40%) suffered losses, problems or costs due to poor quality data. Companies also cited extra costs due to duplicate mailings, excess inventory, inaccurate billing and lost discounts, as well as customer dissatisfaction, delays in deploying new systems and loss of revenue.

Key Data Integration Functions addressed by data integration/ETL software:

In order to promote quality data for use in business intelligence, data integration/ETL software addresses the following key functions:

- **Data Profiling.** Data integration software can be used to automate the profiling of the source data needed for business intelligence. This automated process can help identify out of range values, missing, values, calculated values, and value relationships. Advanced algorithms can be applied to identify problem areas that would be difficult if not impossible to do with manual efforts.
- **Data Extraction.** Data integration software can be used to automate the extraction of data from source systems into the target data warehouse or data mart environment. This can be done in batch or real time mode and scheduled in advance or activated by trigger events.
- **Data Cleansing and Standardization.** Data integration software can be used to standardize the source data (such as filling in missing values or standardizing on street addresses) and cleansing the source data (such as correcting invalid codes).
- **Data Transformation.** Data integration software can also generate corrected data based on prebuilt decision rules.
- **Data Loading.** Data integration software can automate the actual loading process for maximum efficiency.
- **Master Data Management.** Data integration software can be used to automate master data management in a federated manner (where the data stays in the source system) or in a centralized manner (where a master “golden copy” of the data is kept in a central repository). This function is helpful for standardizing references to the same customers, suppliers, or product when these entities are recorded in multiple State application systems. Master data management can also help facilitate applications such as “one stop change of address” where a citizen’s address change can be made in one application system and then automatically updated in all other systems where that citizen’s address is maintained.

Assessment of Current North Dakota Business Intelligence

In preparing the recommendations in this report, CIBER reviewed existing research that had been gathered by the State’s BI Study Team. This section summarizes some of the existing research that was used in the BI assessment.

Inventory of State Business Intelligence Projects

The BI Study Team developed an initial inventory of major State business intelligence application projects. A summary of these current projects are shown below (details are shown in Appendix A):

- **PeopleSoft Data Marts** - the State is licensed for PeopleSoft’s Enterprise Performance Management (EPM) software and five individual data marts (FIN, HRMS, CRM, Supply Chain, and ESA) as well as a number of the “portals.” Included with PeopleSoft is licensing for Ascential ETL tool – which

can only be used to import data for use by the PeopleSoft EPM software. The State has not deployed EPM or any of the data marts in a production environment.

- **PeopleSoft ad-hoc reporting** - the State replicates the PeopleSoft financial and HRMS operational database nightly (MS SQL database) to provide a query environment without impacting the operational database. PS Query is used by power users to generate reports.
- **Game and Fish Management System** - Game and Fish developed a management reporting system using SAS BI software with the help of SAS consultants. The system provides performance measures used by all levels of management in the agency and includes labor reporting, project information, and financial information.
- **Dept. of Health Disease Reporting** - the Dept of Health developed a system to summarize and report disease instances in the state. The system provides visual/graphical trend reporting using SAS.
- **Dept. of Health - Women, Infants and Children (WIC) Program** - The WIC program uses a purchased software application that incorporates Microsoft's BI Analyses Services module that is included as a free component of Microsoft SQL Server.
- **Bank of North Dakota (BND) – Core Banking** - BND is currently in the process of replacing their core banking system with a packaged software solution that uses an integrated version of COGNOS BI software to provide analytics.
- **Job Service North Dakota (JSND) – Labor Market Information - JSND** is currently in the process of procuring a replacement system for its labor market information system (LMI). The LMI system provides a data warehouse and statistical employment reporting internally and via the web using FoxPro software.
- **Workforce Safety and Insurance (WSI) – Fraud detection** WSI is interested in implementing a data warehouse for the purpose of tracking trends in injuries, analyzing rates, etc. They have Oracle tools for multi-dimensional analysis.
- **Department of Human Services (DHS) – Medstat - MMIS** - DHS out sources the Medstat application and transmits data to them monthly. The application is used to detect fraud in Medicare/Medicaid claims. DHS is evaluating the need for a data warehouse and analytics as part of the replacement for their Medicaid Systems. The vendor they have selected, ACS, intends to use Cognos for development of the decision support system and the Informatica data integration (ETL) software to load data into the new MMIS system. Funding for the new MMIS system is up for legislative approval in the next session of the legislature (decision expected by March 2007). .
- **Tax Department GenTax** The Tax Department is in the process of implementing an integrated tax system call GenTax. Integrated within the product is a tool called “Discovery” that identifies potential matches to individuals in both internal and external files.
- **Courts – Data mart** The Judicial Branch uses a data extract tool called Data Mirror to replicate their AS400 data to SQL Server to make the data more accessible to external users.
- **Department of Public Instruction (DPI) – Tetradata** DPI implemented a project to provide data warehouses for school districts across the state and one for DPI. Tetradata (www.tetradata.com) was the software system implemented. Adoption has not been widespread and the project has been abandoned. Several larger districts are exploring alternative solutions. DPI has implemented a statewide student identification checking system to ensure student data is not duplicated in different districts.
- **ND University Systems (NDUS) Performance Measures** – Round table performance measures NDUS implemented a data warehouse (Oracle) several years ago to provide performance measure reporting. The warehouse has been abandoned due to lack of use.

- **FINDET** - FINDET is an interagency cooperative effort to provide follow-up information on North Dakota education and training programs. FINDET is also being used to meet the accountability measure requirements of the State legislature regarding training programs. FINDET uses a technique of data linking to match information on individuals in various educational and work programs to determine outcomes.

Current Business Intelligence Related State Technology Standards

The State currently has several related business intelligence technology standards. These standards are shown below:

NOTE: The following represents current BI related State technology standards:

Report Writer Tools:

- Business Objects Crystal Reports
- Jinfonet JReport
- Oracle Discoverer
- Oracle Reports

Enterprise Databases:

- Oracle
- MS SQL Server
- IBM DB2 UDB

Other Databases:

- MS Access

Assessment of State Business Intelligence Practices

The State BI study team has documented several observations regarding the use of business intelligence applications to date. These observations included:

- The BI approach to date has been one of heavy reliance on consultants to come in, build a data warehouse and leave.
- Most existing BI applications have generally been developed to solve a particular problem or fulfill a specific need for information without regard to the “bigger picture” or enterprise wide impact of the BI applications (e.g. cross agency opportunities).
- Because most BI applications to date have been vendor-driven, little expertise exists within state government surrounding the deployment or support of business intelligence solutions
- The State has had some failed data warehouse projects. In part, this is because the knowledge left with the consultant and there was no ongoing development and support.
- The selection of the BI related toolset on past projects has also been left primarily to the vendor. This has created a hodgepodge of BI technologies to support with no plan for shared information.

BI Assessment Approach

To conduct the assessment of the State's business intelligence needs and directions, CIBER utilized a systematic assessment approach. This section provides additional details on the approach used and resources. This section provides a description of the approach and resources used by CIBER in conducting the BI assessment for the State.

BI Analysis Approach

To address the objectives of the BI Assessment engagement, CIBER utilized a systematic approach in the review of the State's BI environment against BI best practices and experiences in government and non-government settings. This approach involved the following components:

- Review existing documentation regarding the State's business goals, culture, IT environment, and history of BI projects and applications.
- Reviewed high level State BI requirements document prepared by the State BI Study team (see Appendix B).
- Conducted selected on-site interviews with State personnel.
- Analyzed research gathered in the steps above to determine key issues and problems associated with the State's current BI environment and directions.
- Mapped State BI assessment findings against industry best practices and BI implementation experiences from other government and non government organizations.
- Conducted a gap analysis between the current State environment and appropriate best practices and lessons learned from other BI projects.
- Developed specific BI improvement recommendations for the State.
- Developed suggested next steps for implementing the initial BI improvement recommendations.

Key Resources

Key resources used by CIBER in developing the BI recommendations for the State include the following:

- Research reports and input from the Gartner Group (2005, 2006)
- Research reports from the Forrester Group (2005, 2006)
- Selected research from key industry BI related publications (2005, 2006)
- Experiences of CIBER consultants with hands on experience with commercial and state and local government clients – including a business intelligence engagement with the State of Tennessee which received the 2003 Business Intelligence World Class Solution Award from *DM Review* magazine as the best worldwide BI solution for 2003.
- The 2005 OLAP Research Report
- BI Research reports from the Butler Group (2005, 2006)
- *CIO Insight* Vendor Satisfaction Survey 2005
- The Data Warehousing Institute Annual Conference sessions (2004 and 2005)
- IDC Business Intelligence Market Research Report 2005
- Leading Government and business trade publications (2005, 2006).

Key Findings

Based on the limited review of existing documentation and on-site interviews with key State staff, the following key findings were identified:

- **Multiple BI Software Products in Use.** Currently, the State has a large number of different BI software products in use in various agencies. Vendors represented by these various BI software products include: COGNOS, SAS, Oracle, and Microsoft. Most of the BI applications associated with these existing BI products do not represent large BI applications. Continued use of multiple products is seen as promoting “siloes” repositories of key state data that will be difficult to share with other state agencies.
- **Multiple Data Integration (ETL) Software Products in Use or Planned.** The State owns or is considering the purchase of multiple data integration (ETL – extract, transform, load) software products. Vendors represented by these software products include: IBM (formerly Ascential), Informatica, and Data Mirror.
- **No statewide BI coordination or strategy.** There are no statewide strategies or plans for coordinating the development or management of BI applications in the State. This has resulted in several “siloes” pockets of BI applications and the use of multiple BI tools.
- **Limited in-house BI expertise and skills.** Current State BI application users have limited BI expertise as many of the applications were developed through consulting engagements that include limited BI training and knowledge transfer.
- **No formal BI methodologies.** There are no formal BI methodologies being used by the State in the development or operation of BI applications.
- **No formal BI requirements for new application development projects.** There are no BI requirements linked to the development of new application systems or to the purchase of new application software.
- **The existence of a large amount of legacy data (including Adabas and DB2 data).** The state owns and plans to maintain a large number of legacy applications that have data residing in older database formats (e.g. Adabas) that represent source data for future State business intelligence applications. The State’s current BI tools are not adequate for accessing and using legacy data for future business intelligence applications.
- **Business intelligence is underutilized.** The State has significant opportunities to increase the use of business intelligence technology to support critical State decision making. These opportunities cross all state agencies.

Recommendations

Based on CIBER's assessment of the State's BI requirements and directions, twelve recommendations are offered for consideration by the state of North Dakota. Details regarding each recommendation are presented in this section.

1. Enterprise Business Intelligence Software Standard

Based on current and projected State BI applications, CIBER recommends that Cognos be selected as the primary enterprise BI software standard with SAS as a secondary standard to be used in special situations where the Cognos solution may not be sufficient.

Rationale for Selecting Cognos and SAS

Should the State select the MMIS solution recommended by ACS, the State would acquire Cognos as a BI tool for a major State BI application (Medicaid represents a large and important data warehouse application). In addition, the new CORE banking system being implemented by the Bank of North Dakota also includes integrated Cognos BI software. (It is noted however that it is not likely that the Bank of North Dakota would be sharing much information with other State agencies).

Even if the Cognos MMIS solution is not selected by the State, Cognos is recommended as the enterprise BI standard for the state of North Dakota. This recommendation is based on the following key strengths of the Cognos solution for meeting North Dakota's BI needs:

- **Ease of use and usability.** Cognos excels in ease of use and usability which is considered critical for promoting wide spread usage of business intelligence applications at the State.
- **Dashboards and scorecards.** Cognos has a robust solution for developing and using scorecards and dashboards.
- **Integration with FileNet.** Cognos has a proven solution for integration with FileNet.
- **Integration with PeopleSoft.** Cognos has a proven solution for integration with PeopleSoft applications.
- **Analyst rankings.** Cognos is one of only three BI vendors currently ranked by the Butler Group in its top Outperform tier (the other BI vendors are SAS and Oracle). Cognos is also ranked the Gartner Group's top Leaders quadrant for Business Intelligence software suites.
- **SOA integration and flexibility.** Through a strategic alliance with Composite Software, Cognos offers a flexible and robust solution for SOA (Service Oriented Architecture) integration.

While Cognos has many great features, there are areas where the State would benefit by having a second enterprise BI standard in special situations associated with data mining, predictive analytics, and integration with the State's ESRI GIS (geographical information system). The three key functions where SAS may provide significant benefits over Cognos are as follows:

- **ESRI GIS – SAS** is a close alliance partner with ESRI while Cognos is an alliance partner with ESRI's competitor Map Info. Because North Dakota has a large and strong ESRI GIS application,

SAS can help offset this lack of ESRI integration with Cognos (it is noted that Cognos can be made to work with ESRI GIS with some effort and cost).

- Advanced Analytics and Data Mining. SAS excels in this area while Cognos is not as strong in this area.
- Text mining. Again SAS is strong in this area while Cognos is not as strong.

In addition to these benefits, use of SAS as a second standard would allow existing State SAS applications to continue to be used. To resolve the issue of multiple Meta data repositories, both SAS and Cognos software includes the MITI Meta Data Integration software to facilitate the sharing of Meta data between the SAS and Cognos metadata repositories.

Comparative Ratings: SAS vs. Cognos

The chart below represents CIBER’s relative assessment of the SAS BI solution compared to the Cognos solution. The grading scheme uses A for a top (best) rating in the category, B for a mid rating, and C for low rating.

Grading of Business Intelligence Options by Key Requirements

REQUIREMENT	COGNOS grade	SAS grade
Performance	B	A
Ease of use / usability	A	B
Data Integration (ETL)	C	A
Data Quality (ETL)	C	A
SOA Integration	A	B
Cost/Licensing	C	C
Reporting	A	C
BI Strategy	A	B
OLAP analytics	A	A
Data mining	C	A
Text mining	C	A
Data search	A	A
FileNet/BPM integration	A	C
GIS (ESRI) integration	C	A
Corporate Stability	A	A
Mainframe integration (ETL)	C	A
PeopleSoft ERP integration (ETL)	A	B
Predictive Analytics	C	A
Service and support	B	A
Dashboards/Scorecards	A	B
Performance management	A	A
State government vertical applications	A	B

The Other Business Intelligence Software vendors

Other BI software vendors were also considered for possible use at the State. The reasons why these other BI software vendors are not recommended for further consideration are listed below:

- Microsoft SQL Server 2005 – This new release includes both BI and Data integration capabilities. However, both functions are considered light weight and insufficient for a large organization with complex and diverse technology platforms such as the state of North Dakota. While Microsoft’s option is relatively inexpensive on the front end, considerable work would be required to build functions that are standard in Cognos and SAS. This quote from the Gartner Group’s Dec 2005 evaluation of BI vendors underscores weaknesses in the current Microsoft BI solution:

Although Reporting Services is feature-rich, organizations have been reluctant to embrace it because it lacked a self-service query capability until Report Builder with SQL Server 2005. Its licensing and packaging options are the most attractive in the BI platform market. However, it has yet to execute on much of its strategy, such as integration with the SharePoint and Office products.

- Oracle 10G Data Warehouse - This new release is a big improvement over previous Oracle business intelligence offerings. However, the functionality is still incomplete and it is relatively immature being a new offering. In addition, the full BI functionality is deeply embedded into the Oracle data base software which is seen as potentially restrictive and confining – particularly if the State plans to use data appliances – such as Netezza (see recommendation #3).
- Business Objects – This platform is considered similar to Cognos but is not considered as close a fit to North Dakota’s specific BI needs as Cognos. Cognos also has better integration with FileNet which is a major State standard for document management. Also, the Butler Group did not include Business Objects in its top tier “outperform” group for 2006-2007 (only SAS, Cognos, and Oracle are in this group).
- The following vendors were not considered because of their low analyst rankings and because they lacked BI strengths that were considered important for the State: Hyperion, MicroStrategy, Actuate, Proclarity, Information Builders, and Applix.

Using Cognos as the Primary BI Solution

Cognos provides a highly functional and fully integrated BI solution for the State that can meet the needs of all levels of BI users – from power users to “read only” report consumers. Key features in the Cognos BI solution include⁶:

Single interface for each user’s BI needs: Each business user accesses a single, familiar interface delivering consistent and complete information automatically in the format that makes sense to the user role. Based on a person’s role, IT can enable more or less functionality at the push of a button to fit the need.

Single, adaptive authoring environment for all BI capabilities: Gain a single adaptive authoring environment and adaptive report-authoring model for the complete range of enterprise reports,

dashboards, scorecards, and other BI capabilities, based on a combination of either OLAP or relational data sources.

Create dashboards and reports in one product: Create visually engaging dashboards with gauges, charts, maps, and other graphical elements with built-in, fully automated drill-through capabilities—all from the same reporting interface you use for production and business reports. Use all of the same reporting distribution options for your dashboard, such as scheduling and “bursting” (delivering reports to different lists without rerunning the queries for every destination).

Cooperative report development: Various users can collaborate when producing a report. A business user can author a simple ad hoc report and give it to others for their reference, or to a professional author to be enhanced or edited. The report then returns to the business manager, and he or she can share the finished report(s) with intended recipients. Integrated security ensures people only view what they should.

Guided analysis from scorecards, dashboard, or report: View information in a scorecard, dashboard, or report and get further detail all in one product, with content automatically maintained as you navigate. Interactive, dynamic scorecards, dashboards, and reports let business users follow their questions about how the company is progressing against its strategy.

Deep comparative analysis: Analyze large volumes of data from all business dimensions, regardless of where the data is stored—OLAP or dimensionally aware relational sources, or any relational source. Use drag and drop techniques such as pivoting and filtering to analyze large data sets with fast response times. All BI capabilities in Cognos 8 Business Intelligence provide a dimensional view on top of relational data.

BI Search: Cognos Go! Search Service, a new BI search capability, will enable users to instantly find relevant, strategic enterprise information available through Cognos 8 Business Intelligence. With a familiar, browser-based search that ranks results based on user priorities, Cognos Go! will deliver users an intuitive and efficient way of quickly finding BI information and data presented in reports, analyses, dashboards, metric information, and events across the organization.

Microsoft Office Integration: In addition to the ability to embed refreshable reports within Microsoft® Excel®, users can now view, interact with, and refresh Cognos 8 BI reports, analyses, and metrics within Microsoft PowerPoint® and Excel. This ensures centralized control and manageability of information and helps extend the reach of trusted BI content to more users within a familiar environment without requiring additional IT support. Cognos Office Connection enables users to enhance Microsoft Office documents with accurate, rich data, and share those documents and BI content throughout the organization

Event lifecycle management: Event lifecycle management goes beyond the basic notification functionality provided in other products to automate the decision-making process, launch business processes, and integrate with Business Process Management. Where human intervention is required, through decision-process automation, event lifecycle management notifies the people who are accountable and provides the relevant information they need to resolve the issue. Where human intervention is not required, it initiates business processes such as data updates or Web service procedures to resolve the issue. Throughout the event lifecycle, it continues to monitor and notify, ensuring that all relevant players and systems take appropriate action at the right time. Event lifecycle management benefits include:

- Link different parts of the organization together for enterprise-wide event management
- Compress the time to action and resolution.
- Define compound events and tasks using event conditions.
- Manage events in context.
- Use both decision-based and business-process issue resolution.
- Track critical events.

Using SAS for Data Mining and Advanced Analytics

It is projected that in certain situations, the State should consider using SAS instead of Cognos for advanced data mining and text mining applications. In these cases, Cognos' OLAP analytics may not be sufficient.

SAS specializes in data mining and advanced predictive analytics. They hold approximately 36% of the market share in data mining with the 2nd place vendor holding approximately 5% (Source IDC). SAS is also the top rated data mining vendor in the Gartner Group magic quadrant (and one of only two vendors in the top quadrant – SPSS being the other).

SAS particularly excels in supporting the following advanced analytic/data mining functions:

Predictive modeling can be used for fraud detection, failure analysis, predictive maintenance and risk management.

Descriptive modeling can be used to analyze important relationships. An example is analyzing criminal behavior or student behavior to forecast future performance.

Forecasting can be used to project future conditions as a means of planning for optimal staffing levels or ordering critical materials. It can also be used for projecting traffic volumes.

Optimization capabilities can be used to identify the most effective combinations of critical success factors to produce targeted results within known constraints – such as funding, personnel, or equipment.

Text mining can be used to extract structure and meaning from unstructured textual data. By exploring and modeling large amounts of structured data as well as text-based data, organizations can uncover hidden relationships and patterns of information, enhancing their ability to make accurate, on-target predictions.

Experimental design can be used to isolate the specific factors that influence outcomes. This is done by tracking the effects of multiple factors against observed outcomes. This helps to pinpoint which factors have the most impact on target outcomes. This can be particularly helpful in diagnosing the impact of medical treatments, medications, and physical activity on health problems.

Examples of SAS's dominance in this category can be seen in these quotes provided by SAS:

“Based on vendor strategy, prospective buyers of software products for predictive analytics should pick SAS for its uncompromising dedication to data analysis.”

“Vendor Scorecard: Selecting Software for Predictive Analytics” (Lou Agosta, Forrester Research, December 2003)

“What separates SAS’ approach from the other leading business intelligence (BI) vendors is the capacity to understand, and manipulate, statistical processing, a.k.a. analytics.” In the report, Giga recognizes that very few vendors can provide the advanced analytic capabilities that SAS can, and has, for 28 years.

“SAS Brings Something Novel to BI Analytic Applications: Analytics” (Keith Gile, Forrester Research, April 2003)

“SAS has always been known as a stellar provider when it comes to high-end analytic competencies. They have not only enhanced this feature set, they have also carefully tailored their deep analytics into the new SAS®9 platform to create a powerful new form of business intelligence...”

David Stodder, editor-in-chief, *Intelligent Enterprise*

Sharing MetaData between SAS and Cognos

Since both SAS and Cognos maintain separate metadata (data about data) repositories, there will be a need in some cases to share this metadata. Both SAS and Cognos have bundled metadata bridge software from Meta Integration Technology Inc. (MITI) to facilitate the sharing of metadata between repositories (see announcements below from SAS and Cognos).

SAS: The partnership between SAS and Meta Integration Technology, Inc. (MITI) brings added-value and innovative competitive advantage to customers. Managing and sharing metadata with technologies from SAS and MITI is cost effective and low risk. MITI provides the Meta Integration Model Bridge (MIMB) to transport metadata between design tools, repositories, or environments. Integrated with the SAS system, IT can easily manage and share metadata with fewer resources and time to develop complex programs. The ability to import and export metadata leverages metadata for analysis, increases the value for business intelligence and enables better management decisions. SAS and Meta Integration Technology Inc. have established a strategic partnership to provide an enterprise-wide metadata architecture for minimizing risk and cost of best of breed warehousing solutions.

Cognos: September 10th, 2003 -- Cognos and Meta Integration Technology, Inc; today announced a significant Original Equipment Manufacturer (OEM) agreement. Meta Integration's metadata movement components are embedded within the new Cognos ReportNet modeling tool, known as Framework Manager (FM). This new OEM agreement provides Cognos ReportNet with an integrated metadata import/export solution for popular third party modeling tools, as well as the OMG CWM.metadata standard.

2. Enterprise Data Integration/ETL Software

While Cognos includes some basic ETL functionality, it is not seen as sufficient for the long term – primarily because of weakness in linking to legacy application data bases without the use of third party software (e.g. Adabas, DB2, IMS, etc.). Consequently, it is recommended that the State consider the purchase of either the SAS Data Integration software or the Informatica Data Integration software. It is further recommended that this decision not be made until it is determined whether the State will be implementing the MMIS solution that includes the Informatica software (decision expected in March 2007). If the State does pick the MMIS solution that would come with Informatica, it is recommended that the State select Informatica as the state standard. If the State does not pick the MMIS Informatica solution, it is recommended that the State conduct a comparative evaluation between SAS and Informatica at that time.

SAS is recommended as an option particularly if the State is using SAS as the secondary BI software standard (as recommended in recommendation #1 above). SAS is recommended because it has native access to legacy data sources (Adabas, DB2, and IMS) and is highly rated by industry analysts. It is also one of the top three data integration software vendors that control 60% of the data integration software marketplace. Informatica is recommended because it is the market leader, is less expensive than the other major data integration software vendor (IBM's Websphere Data Integration software (formerly Ascential Software's Data Stage, Quality Stage and Profile Stage products).

State government data integration/ETL case study

As an example of how the SAS or Informatica Data Integration software can benefit the state of North Dakota, information regarding an actual state government case study is shown below. The information for this case study was obtained from Informatica's public web site. In this case study, the state of Washington Department of Transportation used Informatica's data integration/ETL software to access legacy Adabas information and will eventually use Informatica to display reports with ESRI's GIS system.

The Challenge

Throughout the 20th century, the Washington State Department of Transportation engineered a highway infrastructure that traverses more than 7,000 miles of road and 3,300 bridges, from the wheat fields of Eastern Washington to the rugged coastline of the Pacific Ocean. Washington's transportation system is recognized as one of the most accessible, efficient, and safe systems in the United States.

Until several years ago, however, the same could not be said of the IT infrastructure that supported the DOT's financial, construction, maintenance, and traffic safety programs. Most of the agency's information systems—some dating to the late 1970s—are based on an IBM mainframe and Software AG's Adabas database, with a highly indexed, hierarchical data structure at odds with modern relational data systems.

Leveraging legacy data from the insular Adabas environment demanded tedious custom coding, generation of flat files for transformation into relational formats, and high consumption of mainframe computing cycles. Manual development of extraction utilities was costly, inefficient, inflexible to changing requirements, and difficult to maintain by anyone other than the original programmer.

Moreover, this complex, brittle IT infrastructure inhibited access to data that the DOT recognized could be analyzed for insights into improving traffic and highway safety, finance and accounting across multiple divisions, and metrics-driven management of nearly \$300 million a year in capital projects and thousands of contracting firms. To get information into the hands of business-side analysts, the DOT embarked on a data warehousing initiative. The department needed a data integration platform that could:

- Natively access specialized mainframe and Adabas data
- Eliminate customized, ad hoc data access mechanisms
- Supply flexible, high performance data extraction, transformation, and loading
- Provide investment protection through standardized data integration

The Solution

After proof-of-concept testing of the two vendors that offered native access to Adabas data, the DOT in 2001 selected Informatica PowerCenter and PowerExchange for their superior reach into hierarchical Adabas data, simple drag-and-drop development GUI, and the promise to mature into rock-solid technology that would deliver sustained investment protection over many years.

According to Chris Kemp, the DOT's manager of Data Management Services, the Informatica platform in the ensuing four years has more than delivered on promise. "A big benefit is that Informatica gives our developers confidence that the data integration processes they're coding today are going to be around and maintainable. We see it as a product that will last a number of years-we'll be able to take advantage of the skills we're building into it for a long time."

As of late 2005, the DOT has used PowerCenter to build and maintain eight data marts, based on a Microsoft SQL Server database with a Web-based Hyperion business intelligence front-end used by roughly 300 DOT managers and analysts. PowerCenter's pre-built connectivity to Adabas and its ease of use has helped the DOT's Data Management Services, with two skilled Informatica developers, maintain a brisk pace of deploying about two new data marts each year.

Kemp estimated that PowerCenter has enabled DOT to reduce the development time required to build Adabas data mappings and transformations by 25 to 50 percent, compared to previous custom-coding techniques. Moreover, the DOT was able to redeploy six full-time employees who were previously dedicated to data extraction and report generation to more strategic IT initiatives.

Streamlined data integration was instrumental in enabling DOT to rapidly realize its primary objective-improved analysis, insight, decision-making, and management of tax dollars by highway planners, construction engineers, project managers, and other DOT personnel. Its first data mart,

the Capital Program Management System (CPMS), has been vital in helping DOT fine-tune management and budgeting of more than 500 projects in a six-year pipeline.

Other data marts consolidate collision data and support analysis of road conditions, traffic volumes, and other accident factors. It is a key resource in DOT's overall efforts to improve highway safety. In 2004, there were 569 fatalities on state-owned roads—the fewest since 1961—even though the number of registered vehicles had grown to more than to 5.6 million.

Other data marts are geared for accounting, consumable inventory such as road salt and sand, roadway characteristics, DOT facilities, traffic patterns and volumes, and project scheduling. Future data marts will cover employee labor, human resources, real estate, construction, equipment, and budgeting. In all, the warehousing environment includes about 204 GB of data.

The Results

Reduction of 25 to 50 Percent in Development Time and Improved Systems Performance

Since its deployment in 2001, PowerCenter has helped DOT's Data Management Services developers reduce the time otherwise required for custom coding by 25 to 50 percent and re-deploy six individuals previously tasked with manual data extracts and report generation. Those benefits will extend over a number of years as the DOT broadens PowerCenter's role as the agency standard for data integration and eliminates the need for costly resources with specialization in Adabas data access. With an upgrade to PowerCenter 7.2, the DOT has found a substantial reduction in load processing times, so that nightly movement of 2.5 GB is accomplished in less than half the four hours allotted.

Native Access to Mainframe Legacy Data to Drive Analysis and Insight

PowerCenter has provided the DOT with the data extraction and integration component essential to enable it to rapidly roll out data marts that some 300 business analysts rely on for cost-effective project management, budgeting and financial analysis, improved highway safety, and reduction in accidents, injuries, and fatalities. The data marts have been embraced by analysts who previously struggled with mainframe reporting and exporting information to Excel spreadsheets for analysis.

Long-Term Investment Protection through Standardized Data Integration

DOT officials view the agency's deployment of PowerCenter as a sound long-term investment that will pay dividends for years to come by supplying a standardized data integration environment, as well as enabling it to protect and leverage its existing investments in mission-critical legacy data housed in its IBM mainframe and Adabas databases. The DOT also gives Informatica high marks for technical support that has helped it to maximize the value of its data integration platform.

Flexibility for Continued Data-Driven Modernization

With enhanced functionality for data integration in a service-oriented architecture, or SOA, Informatica PowerCenter is likely to play a key role if the agency embarks on an initiative to modernize its 12 key Adabas-based applications into an SOA using Web services standards such as SOAP and WSDL. (In December 2005, the project was under assessment and pending legislative funding approval.) Also on the DOT's roadmap is replacing hand-coded system

interfaces with Informatica mappings, and leveraging PowerCenter to transform Adabas data into a spatial format for use by ESRI GIS (geographic information system) mapping applications.

3. Enterprise Data Warehouse Platform Standard

To house data to be used for BI analysis, it is recommended that the State standardize on a data warehouse platform. Based on an assessment of the current and planned BI applications it is projected that the State will need to be able to support a two terabyte data warehouse environment. To meet this need, it is recommended that the State consider the use of the Netezza data appliance as the primary data warehouse platform for future BI applications. In those cases where a data appliance may not be practical, the use of the Oracle 10G database is recommended as the secondary data warehouse repository standard for the state.

Actual users have found that the Netezza NPS® data warehouse appliance can potentially provide 10-100 times the performance at half the cost of traditional warehouse database systems from IBM, Oracle, Teradata, and others.

To validate this claim it is recommended that the State conduct a proof of concept evaluation test with Netezza, and compare the results against using an Oracle 10G database data warehouse environment.

The Benefits of a Data Appliance for Data Warehouse Applications

Data appliances – such as that offered by Netezza – are designed from the ground up to support data warehouse/business intelligence applications. This design provides inherent benefits over the use of grid based general purpose database solutions such as those offered by Oracle and IBM.

With a general purpose grid structure, data must be moved from storage to server memory before any processing can be done on it. The result is a significant input/output bottleneck as large numbers of rows are transferred back and forth across the grid. In this scenario, general purpose database server grids add very little performance value for complex queries – particularly iterative queries - because the basic processing approach is still the same.

On the other hand, data appliances – such as those from Netezza – are designed to optimize query performance for massive data sets. This is possible because the Netezza solution uses an architecture that moves processing to where the data is physically stored, eliminating unnecessary data movement.

One of the major concerns about the use of data appliances is the lack of data integration with business intelligence or data integration/ETL software. To offset this issue, Netezza and formed strategic partnerships with key software vendors to certify integration with the Netezza appliance. All of the recommended BI and data integration/software products for North Dakota have been certified to work with Netezza: SAS, Cognos, and Informatica).

Another benefit of using a data appliance is that all necessary data storage is included along with automatic back up and recovery. Also, the data appliances can be scaled to handle larger data warehouse volumes by simply adding additional processors.

The Netezza Performance Server NPS 5200 Data Appliance Details

The recommended Netezza product for the State is the Netezza Performance Server NPS 5200. The Netezza Performance Server (NPS[®]) data warehouse appliance architecturally integrates server, storage and database in one easy-to-use system. The NPS 5200 system is a half-rack version of the enterprise-class NPS family of data warehouse appliances and is designed to handle terascale analyses of data, cost-effectively and in an office-area environment.

Designed for high-performance analyses on data sizes ranging from several hundred gigabytes up to three terabytes, the NPS 5200 is an economical data warehouse alternative for medium sized businesses that can't afford the high cost of traditional options. And because the NPS 5200 is small and uses standard power, it can plug into any wall outlet in an office copy room or supply room.

Netezza Software functionality

High-Performance SQL Engine — Optimizes SQL on the NPS system's massively parallel streaming architecture

Database Operating System — Executes highly efficient SQL through intelligent distribution across host and storage-side processing

Fast Loading/Unloading — Allows new data to be added rapidly (up to 500 GB/hour) so that ongoing processing can work with real-time data. Integrates externally with leading ETL and EAI tools.

Supports "trickle feed" real-time data loading concurrent with query processing.

Administration Tools — Includes DBA and Systems Management tools that provide easy-to-use GUIs & CLIs for permissions, monitoring, diagnostics, trouble-shooting and other administrative functions.

Netezza Industry compatibility

Supports full ANSI transactions

ANSI SQL-92 compatibility and SQL-99 analytics extensions

ODBC and JDBC Type 4 API compliant

Windows, Linux, HP-UX, AIX, and Solaris operating systems

Integrates with leading BI and ETL platforms (including SAS, Cognos, and Informatica)

4. Integrate BI with the State's Enterprise Architecture

CIBER recommends that the State's BI standards be integrated as part of the State's overall Enterprise Architecture strategy for data collection, sharing, and reporting. The proposed standard BI and data integration/ETL software should be considered as an integral part of the total enterprise computing strategy and infrastructure for the State. This would include showing how the State BI and ETL software would integrate with current State technology such as the FileNet document management system, the state Active Directory standard, the PeopleSoft ERP system, and the ESRI GIS system. A graphic description of how the proposed State BI Architecture could potentially fit into the State's Enterprise Architecture is shown in Appendix C.

Additional details regarding key potential enterprise integration opportunities are shown below:

1. Microsoft Active Directory Integration

Cognos SOA architecture is designed to snap into other application technology and has been integrated with Microsoft Active Directory frequently. Cognos uses native active directory as an authentication source. Sign-on is configured with a combination of either Apache or Microsoft IIS web application services.

SAS also supports the Microsoft Active Directory as one of many standards supported by SAS. These supported standards include: Lightweight Directory Access Protocol (LDAP), including Sun ONE Directory Server, Netscape Directory Server, IBM Secure Way Directory Server and Microsoft Active Directory implementations. If the SAS Open Metadata Architecture is used, there is a choice to use an LDAP or Microsoft Active Directory server as an authentication provider for SAS Web applications, SAS Metadata Servers and SAS OLAP Servers.

2. Service Oriented Architecture (SOA) Integration

Cognos has a strategic alliance with Composite Software to facilitate SOA integration. Details regarding this relationship can be found in excerpts from this Cognos press release:

BURLINGTON, MA, December 7, 2004— Cognos announced it is embedding the Composite Information Server from Composite Software, the leading independent provider of Enterprise Information Integration (EII), into Cognos ReportNet through an exclusive BI OEM agreement. Cognos has further solidified the partnership by making a \$4.5 million minority equity investment in Composite Software.

With this agreement, Cognos allows new styles of reporting and analysis by combining historic data from warehouses with real-time operational information from a wide range of data sources including XML, JDBC, WSDL and LDAP — enabling a "single version of the truth" — without moving or replicating data. Organizations can dynamically access and leverage all information, internal or external, no matter where the information resides in order to be more competitive.

"Our research into the use of business intelligence applications for enterprise operations shows that integrating data from multiple data sources is a key requirement. Augmenting the data warehouse with dynamic data from other sources that include Java applications, the Internet, and other similar sources drive this requirement," said Eric Rogge, Vice President and Research Director at Ventana Research. "Solutions such as the integration of Cognos ReportNet with the Composite Information Server are a natural extension of BI technology and are a necessary step forward to meet emerging enterprise information needs."

3. ESRI Geographical Information System (GIS) Integration

SAS has a strategic alliance with ESRI to integrate SAS BI with the ESRI GIS system. These excerpts from a recent press release between SAS and ESRI provides additional details regarding current and future integration functions between SAS and ESRI software.

Redlands, CA and PHILADELPHIA (April 11, 2005) – At the 30th annual SAS® Users Group International (SUGI) conference in Philadelphia, SAS, the leader in business intelligence and analytics, unveiled integrated business intelligence (BI) applications that will allow organizations to visualize and analyze data relationships in new ways by considering both location and traditional business information dimensions together.

GIS combines layers of information about a place to provide a better understanding of that place; the layers of information you combine depends on your purpose: finding the best location for a new store, analyzing environmental damage, viewing similar crimes in a city to detect a pattern, and so on. For example, an insurer can combine a layer of existing insured properties with a layer showing FEMA's flood zone boundaries to determine financial impact in the event of a flood. If a disaster actually occurs, a map can be added to the same reports and distributed to field agents who can then more easily locate potentially damaged properties.

The combination of SAS and ESRI will enable organizations to apply the best technology to solve high-value business problems that inherently consider the impact of spatial relationships as part of the intelligence-creation process. Organizations will be hard-pressed to find alternate solutions that provide the power and flexibility that SAS and ESRI provide.

Both ESRI and SAS have designed their latest software releases to integrate BI and GIS functionality into a Web-based, zero-download application running in a browser. This new application allows users to analyze and visualize complex BI data relationships in the context of locational influences. Users will access the new functionality through the SAS Web OLAP Viewer, one of the analysis and reporting interfaces to SAS Enterprise BI Server, and now ESRI's ArcGIS Server.

Users will have access to SAS data analysis and reporting functions as well as ESRI location analysis and map display functions. This integration of the SAS and ESRI platforms allows companies to analyze, display, and distribute traditional business data and

location-based data to multiple departments across the enterprise from a single, integrated platform. SAS and GIS functions will also be available to developers for creating focused, custom marketing, sales, logistics, risk analysis and other applications that blend seamlessly into their existing desktop, browser, dashboard or portal environments.

“The ESRI partnership with SAS has succeeded in making geographic analysis an integral part of everyday business intelligence processes in an environment familiar to both business and government,” said Jack Dangermond, ESRI founder and president. “We are extremely pleased that SAS has chosen ESRI to provide the geographic analysis behind its breakthrough application.”

Advantages of integrating ESRI GIS with the SAS business intelligence platform are:

- Ability to explore and visualize GIS and BI relationships in a familiar Web-based analysis and reporting environment.
- Obtaining more rigorous and useful business intelligence from existing data.
- Communicating discovered intelligence more intuitively through maps.

From SAS Bridge for ESRI brochure

Major features of SAS Bridge for ESRI include:

- Simplified interface for SAS and ESRI communication.
- Read/write attribute exchange to SAS datasets.
- ArcGIS surfacing of SAS metadata.
- Execution of SAS programs from ArcGIS environment is enabled and examples of stored processes are provided.
- Access to your distributed SAS systems and data.
- InstallShield Wizard with checks for SAS and ESRI software.

With SAS Bridge for ESRI, you can increase your efficiency by easily linking tabular data in SAS with spatial data in the ArcGIS environment. This means that queries from ArcGIS can inherently consider attributes developed within SAS. No longer will you need to re-write customized programs every time your source data changes — just link the relevant tables for complete information.

Having access to richer data means greater understanding, and ultimately, better decisions. You can also communicate results better by showing industrial-strength analytical and business intelligence graphics on your maps for immediate recognition and understanding. Since SAS Bridge for ESRI utilizes the ArcGIS interface, users have the industry-leading GIS interface in which to access the power and quality of SAS. In addition, this solution enables you to extend your analytical capabilities to obtain results that inherently consider spatial attributes. You’ll be able to distinguish significant from insignificant findings — going beyond physical coincidence to true, measurable results.

You can utilize SAS to create accurate representations of current and future events, and then treat the results as spatial-references attributes. These attributes can then be visualized

as a thematic layer, and/or used as input for additional spatial analysis. You can extend your use of ESRI's ArcGIS by ensuring that spatial proximities, adjacencies and the like become variables in your analysis — inherently considering the impact of spatial phenomena in your decisions.

Cognos and ESRI Integration

Cognos does not have a strategic alliance with ESRI (Cognos has an alliance with one of ESRI's competitors – MapInfo). Consequently, additional effort will be required to integrate Cognos with ESRI. This can be done by writing programs in Java as a service offering to issue XML data to an ESRI archIMS or archGIS ESRI server. Cognos can plot ERIS mapped data on a Cognos report. Cognos can also offer ESRI functions like zoom, change layers, drill through points on a map. The City of Minneapolis is a good example of where Cognos is integrated with ESRI.

4. FileNet Integration

Cognos has a strategic alliance with FileNet to integrate Cognos BI functions with the FileNet system. This recent press release provides additional details regarding current and planned integration functions.

November 8, 2005- FileNet Corporation and Cognos today announced the two companies have extended their partnership to deliver demonstrable process-enabling business performance management, geared to helping organizations make faster and more effective decisions at the most relevant points in a business process.

Under terms of the extended agreement, FileNet and Cognos will co-operatively market their respective platforms, the FileNet P8 platform and Cognos 8 Business Intelligence (BI), to customers. The two companies will now be able to deliver jointly-developed solution templates that create frameworks for easy application deployments, and provide a holistic view of business process and business intelligence information through a number of integration points:

- Business intelligence on process – process data provided by FileNet is enriched and aggregated with enterprise data from Cognos 8 BI.
- Business intelligence embedded within process – BI information essential for decision-making is easily incorporated into each process step of a FileNet BPM process, so users have all the information required to make the best decision at the right time.
- Event-driven process – proactive monitoring, detection, and alerting capabilities found in Cognos 8 BI's unique business event management functionality enable management-level business processes to be automatically initiated and resolved through FileNet when required.

According to a July 2005 Forrester Research report*, the market is on the brink of what Forrester terms a "Process to Data (P2D)" explosion, as more BI and BPM vendors begin to recognize the business value in establishing formal technology relationships. "What is

different today is that companies are looking for a better way of linking the decision-making process with existing workflow to gain an improvement in business performance,” stated Keith Gile, principal analyst at Forrester Research. “For this to happen, BI/xPM solutions must include a consistent mechanism for defining and using pre-defined processes, while BPM must include a mechanism to analyze and report on business process performance.”

For customers, an integrated FileNet-Cognos offering improves productivity by delivering relevant information at the point in the business process where key decisions must be made. Making business performance data useful to a wide range of employees helps strengthen decision-making across the enterprise, provides greater insight into how work is accomplished, and enables companies to close the gaps between current and desired operational performance goals. Furthermore, organizations can quickly detect a performance trend or issue, and then automatically launch a business process in response to this alert.

“Cognos and FileNet’s commitment to delivering business intelligence embedded in the process holds significant promise for forward-looking organizations like SunTrust,” said James “Bubba” Small, senior vice-president, senior delivery manager at SunTrust Banks, Inc. “As one of the largest commercial banking organizations in the United States, it is fundamental to our ongoing success that we arm our knowledge workers with up-to-date business intelligence data when it matters most, during key decision-intensive points of the business process.”

Additionally, Cognos and FileNet’s process-enabling BI collaboration helps facilitate regulatory and corporate governance compliance. All BI information used in a process can be recorded, facilitating future analysis and providing proper context in which a decision is made. This historical visibility allows customers to prove and substantiate what drove business process decisions, thus delivering critical accountability and audit.

“Enterprises continue to make investments in both automating their business processes, and analyzing their enterprise data. The next logical step is to synchronize the value of both initiatives,” said Leah MacMillan, vice-president of product marketing at Cognos. “Leveraging BI with business process capabilities that span both people and systems is the ignition switch to more active and effective decision-making. By collaborating with FileNet, we are helping organizations gain deeper contextual intelligence about their business processes, which can result in a clearer view of the overall health of a project, a program or the enterprise as a whole.”

“Organizations today must be more agile in responding to customer demands and changing competitive dynamics,” said Martyn Christian, chief marketing officer at FileNet. “To accomplish this, businesses are demanding that information required to make decisions and monitor business performance is directly integrated with the business processes that drive the organization. FileNet’s integration with Cognos creates a unique solution, which other BPM and BI vendors alone cannot provide.”

* -- Source: "Business Process Management (BPM) Meets BI and Performance Management (xPM) Head On" Keith Gile, Connie Moore with Colin Teubner, Forrester Research, 27 July 2005

5. Replace Non-Standard BI Software

CIBER recommends that the State eventually replace current non standard BI related software with the new proposed State standard BI and data integration/ETL software. This would include removing the PeopleSoft EPM Ascential software (see recommendation #8 below) and the Business Objects Crystal Reports software when it is practical to do so. Removing non State standard BI and ETL software will reduce software and support costs, improve the quality of data sharing, and help promote "one version of the truth."

Specific existing software that could be candidates for replacement in whole or in part are:

- Business Object's Crystal Reports software
- Oracle Discover
- PeopleSoft EPN's IBM/Ascential Data Stage ETL software

It is noted that because Microsoft's business intelligence and data integration capabilities are bundled into their SQL database offering, it would be acceptable to use these BI related functions in selected circumstances where it makes sense. For example, the use of the Microsoft BI functionality might make sense for Microsoft SQL server applications where some limited BI functionality is needed but would not require extensive sharing data with other State agencies.

6. Perform BI Assessments of New Application Systems

CIBER recommends that the State perform BI impact assessments for all new application systems developed or purchased by the State. These assessments would identify how the proposed new application systems might benefit from the State's BI and ETL software standards. This process will make it easier to integrate new application systems with BI software when it is appropriate to do so. This process will also help minimize the use of non standard BI and/or ETL software.

The process and policy for conducting a BI assessment for new State applications should be incorporated into the State's IT policy. The responsibility for conducting the BI assessment should be assigned to the BI Competency Center staff (see Recommendation # 7 below).

7. Establish a BI Competency Center

CIBER recommends that the State establish a central BI Competency Center to minimize costs and maximize quality in the deployment and support of BI applications. The Competency Center would work with vendors, outside consultants, and state employees as needed to provide coordination and promote

best practices. An associated statewide BI User Group is also recommended along with a BI User web site for posting useful BI tips, FAQs, and lessons learned.

1. BI Competency Center Functions

Business Intelligence Strategy and Governance

This function involves:

- Developing and maintaining policies, strategies, and BI best practices
- Developing a BI Methodology
- Establishing and tracking performance metrics for the BI operations
- Conducting BI research and attending educational sessions to help the State keep abreast of new BI trends.

Technical Support

This function involves providing first and second level technical support for BI and data integration/ETL software. This should include establishing an internal State web site to post tips and FAQs. This function will also involve interfacing with the BI and ETL vendor technical support groups.

Training

This function involves finding and coordinating appropriate training opportunities for the State's BI users.

Data Stewardship

This function involves assigning and working with data stewards for data maintained in the BI applications. This will also include establishing appropriate security and privacy strategies, policies, and procedures and standards for data quality.

Vendor Administration

This function involves managing BI and ETL vendor contracts and communications.

Software Administration

This function involves serving as administrators of the BI and data integration/ETL software.

2. BI Competency Center Staffing

To staff the BI Competency Center, it is initially recommended that four to five state employees be assigned on a part time basis to the BI Competency Center. Later these positions could become full time depending on how many new BI applications are developed at the State. Potential positions include the following:

An example of how these employees would be allocated is as follows:

- Center Director – oversees strategic plans, directions, and operations for the Center
- BI Administrator – serves as administrator for the Cognos and SAS BI software
- Data Integration Administrator – serves as administrator for the Cognos data integration/ETL component and eventually the State standard data integration/ETL software (Informatica or SAS)

- BI Business Analysis – serves as educator and trainer for business analysts in other agencies on how to use BI. Also would work with consultants to develop a BI methodology for the State.
- Database Administrator (part time only) – this person would assist in addressing data modeling issues associated with building the data warehouse or data marts.

It is recommended that the State in conjunction with outside consultants, incorporate a BI Methodology that would include policies, procedures, best practices, tips, etc to guide the development and operations of BI applications. It is further recommended that this BI Methodology adopt principles from the Capability Maturity Model (CMM) for data warehousing projects.

Capability Maturity Model (CMM)

It is recommended that the State’s BI Competency Center incorporate the Capability Maturity Model (CMM) to provide governance for the development and management of business intelligence applications. The CMM represents a progressive international benchmark of business practices and standards that have been universally accepted as critical for the ultimate success of a project or venture. While there are many different applications of the CMM model, the model most relevant for information technology (IT) endeavors is the Systems Engineering Capability Maturity Model. The Software Engineering Institute (SEI) of Carnegie Mellon University, the Department of Defense, and other entities developed this model.

The CMM is designed to be an easy-to-understand methodology for ranking the use of quality processes related to the development and management of information technology within an organization. The CMM has six levels, 0 through 5 (see Figure 1):

- Level 0 Not Performed
- Level 1 Performed Informally
- Level 2 Planned and Tracked
- Level 3 Well-Defined
- Level 4 Quantitatively Controlled
- Level 5 Continuously Improving

The purpose of these levels is to provide a "measuring stick" for organizations looking to improve their business intelligence application development processes.

There are several benefits for using the CMM model for developing business intelligence applications:

- It provides an easy to understand framework for both IT professionals and those that do not have a background in IT
- The CMM is useful for understanding and viewing major IT issues/concepts.
- The model allows for continuous improvement progress and progressive improvement. It is intuitive that a company cannot currently be ranked at a level 2 and jump
- The CMM provides metrics that can be used as a performance “scorecard” for evaluating the effectiveness of the business intelligence applications and infrastructure.

The CMM consists of five key concepts: consistency, repeatability, transferability, quantity, and quality. Specific examples are as follows:

- Consistency indicates that a key required task is performed much more frequently than it is not performed.
- Repeatability refers to activities that provide proven value and can be replicated by other BI project teams

- Transferability means that the BI best practice function is standardized and followed throughout the company. As a result, the success of this task is transferable across groups within the company.
- Quantity refers to the ability to measure output of an IT activity (such as time and cost)
- Quality refers to how well a task was accomplished (such as meeting the customer’s decision making requirements).

8. Consider using Cognos BI for the Peoplesoft BI applications

While the State owns the Peoplesoft EPM software, it is recommended that the State consider using the Cognos BI tool for the Peoplesoft BI applications. Use of Cognos BI tools will make it easier to integrate Peoplesoft data with non Peoplesoft application data. In addition, since the State uses the Microsoft SQL database instead of the Oracle database for the Peoplesoft applications, it is believed that future enhancements to the Peoplesoft MS SQL EPM database versions will lag enhancements in the Oracle database versions. It is also projected that Oracle may favor its own Performance Management software for future enhancements over the Peoplesoft EPM software based on this information obtained from the Gartner Group.

“Gartner believes Oracle’s own EPM offering will dominate. . . Oracle application customers that own JD Edwards applications should consider Oracle’s EPM rather than the Oracle Peoplesoft EPM offering,”

Gartner Group analyst, 2006

To validate this potential use of Cognos BI for the Peoplesoft Bi applications, a proof of concept test is recommended.

Cognos – PeopleSoft Integration details (from Cognos literature, Sept 2005)

Cognos 8 BI is built to extend existing capabilities in a variety of PeopleSoft solution areas. Whether invested in an application such as PeopleSoft Human Capital Management (HCM), Financial Management, Customer Relationship Management, Supply Chain Management, or the PeopleSoft Enterprise Performance Management (EPM) platform, Cognos 8 BI can add value with a single solution that includes reporting, analysis, scorecarding, and event management.

Cognos works with over 2,600 PeopleSoft customers to provide comprehensive BI and CPM solutions. Cognos 8 Business Intelligence provides all BI capabilities on a single, proven Web services architecture that integrates readily with PeopleSoft solutions.

For PeopleSoft customers, Cognos provides three key benefits:

- Automatic use of PeopleSoft business content and security.
- A single source for BI analysis, reporting, scorecarding, and event management.
- Consistent business visibility across multiple business applications.

COGNOS 8 BI and PeopleSoft EPM

Cognos 8 BI provides timely and complete reporting, analysis, scorecarding, and event management to leverage your existing PeopleSoft EPM investment. Cognos 8 BI and PeopleSoft EPM combine data from

multiple sources, including non-PeopleSoft data. The data integration capability in Cognos 8 BI reuses your existing PeopleSoft EPM security and metadata and makes it available to all other Cognos 8 BI capabilities.

PeopleTools Cube Manager

As a PeopleSoft customer you have access to PeopleTools Cube Manager. Cube Manager speeds up the process of creating OLAP cubes for use with Cognos Analysis. Cube Manager leverages PeopleSoft Query and Tree Manager, which enables the bypass of the manual steps typically required building and updating OLAP cubes.

9. Consider BI Cost Sharing with Other Organizations

While the focus of this BI assessment was on the State of North Dakota, it is recognized that other State organizations could also benefit from sharing the proposed State BI resources (hardware, software, and competency center). Consequently, it is recommended that the State consider sharing some of the costs of the BI Competency Center, the Data Warehouse platform (Netezza) and the BI and data integration/ETL software with other organizations – such as - Bank of North Dakota, the North Dakota University system, and local governments and school districts. In addition, some of the costs of the proposed State BI resources may be eligible for reimbursement with Federal funds in those cases where State BI resources are used to support federally funded programs.

How Universities and Colleges can benefit from Cognos and SAS

The information below has been extracted from Cognos and SAS case studies to show how universities and colleges in North Dakota could potentially benefit from using Cognos and SAS BI software.

University of Minnesota's use of Cognos

(Originally reported in *Campus Technology Magazine*) After implementing an ERP system from Peoplesoft - now owned by Oracle - the school needed a business intelligence system to organize and present its mountain of data in meaningful, actionable ways. BI deployment highlights included:

- Almost immediately, school officials found and recovered \$5 million in unbilled tuition.
- Using different data cubes, the university distributes gift aid more strategically, identifies and counsels students with academic troubles, and has improved its management of class sizes.

Penn State's Use of Cognos

Penn State employs Cognos business intelligence on its Enterprise Information System (EIS) in a number of areas, including alumni, development, the Graduate School, human resources, enrollment, and admissions. The system offers access to these subject areas using a strategy that makes accessible general models for all EIS users and additional secure models restricted to designated users only.

Once Penn State implemented Cognos, users could access the EIS system to easily answer questions and retrieve data and reports. In the past, they would have had to query a data warehouse or a mainframe database, or would have needed to call someone else, to find the answers to their questions. As in most organizations, this resulted in undue pressure on time and resources and unnecessary delays.

Now users can retrieve ‘quick counts’ almost instantly. For example, an alumni model provides counts on the members of the Alumni Association, as well as data on a wide variety of subjects in this area: when alumni graduated from Penn State, which academic unit they graduated from, their age ranges, ethnicity, gender, and so on.

The same approach can be applied to enrollments. Users can look at student enrollment by ethnicity, age range, ranges of grade point averages, by college of enrollment, and so on. This is not detailed information, but high level counts of various populations of students. Users can then slice and dice the data according to any number of different factors. The same approach is also being used in HR—users can get quick counts of numbers of employees by gender, by ethnicity, by date of hire, and so on. This is extremely useful and helpful for HR managers. Users can also analyze the information to track and reveal trends. To take one recent example, users have been able to analyze the states from which the greatest number of Penn State applicants have come in recent years. Upon identifying the top two states (other than Pennsylvania), the University has focused its recruiting efforts, and now employs full time recruiters in those states. This lowers recruitment costs and increases both the number and the quality of candidates.

The Graduate School is currently using Cognos to conduct a major internal program review of 147 graduate programs. Trend over time data is pulled from a number of areas: admissions, enrollment, new student data, average GRE scores, average MAT, average G-MAT scores, average time to degree, and so on. This information is then sent to members of a Graduate Council committee, which reviews the data and additional information provided by the graduate programs and identifies areas for improvement. In many instances, this type of analysis leads to new understandings of the process. For example, several programs discovered that their average time to degree was longer than expected. This underscores a core value of the Cognos solution that allows the organization to move from an anecdotal view of the business to a fact-based view of the business. This enables better decision making and leads the way to a better performing organization.

Cognos is also used by the Graduate School for compliance reporting, to report back to state and federal organizations that help fund university programs. For example, as a part of grant applications prepared by faculty, the National Science Foundation requires data on the School’s programs, as well as demographic information that includes gender and minority populations.

University of Alabama Birmingham’s use of SAS

CARY, N.C. (Aug. 10, 2005) – SAS, the leader in business intelligence, today announced that the company will team with the University of Alabama and the University of Alabama at Birmingham (UAB) Health System in an effort to improve patient care. SAS is sponsoring a pilot study that

will be conducted by University of Alabama MBA students, who will use SAS® data and text mining solutions to examine 12 thousand of the 7 million records in UAB Health System’s patient database.

UAB Health System selected SAS Enterprise Miner™ and SAS Text Miner to analyze information from doctors’ notes and textual comments in its patient database. SAS delivers advanced analytic technology that will convert unstructured data into intelligence that can be mined in concert with demographic variables to track and predict patient response to various medical procedures. The UAB Health System is looking for patterns and clusters of repeated indicators to point to the best methods of treatment.

Partnering with institutions of higher education is nothing new for SAS. In the past four years, the privately held software company has worked with more than 100 universities and colleges to provide trainer kits, guest lecturers for MBA courses, and innumerable workshops and conferences. The University of Alabama’s MBA students will be given an opportunity to work with patient records from UAB Health System to solve real-world issues using SAS software. Hardin reports that students who have completed similar predictive modeling studies supported by SAS have landed jobs after graduation with annual salaries \$10,000 to \$15,000 above their peers.

And for SAS, it just makes good business sense. “This partnership gives students additional tools to survive in the competitive work force and helps address pressing healthcare issues,” Oglesby said. “It also raises awareness of the power of SAS software in college institutions that are training future business leaders. This is a situation where everyone wins.”

10. Establish a Statewide BI Application Advisory Council

The State should consider the establishment of a permanent Statewide BI Application Advisory council of senior State officials to identify and prioritize the development of business intelligence applications for the state of North Dakota. This council could also be a subset of existing State IT advisory committees or groups. The Council should include elected officials and the State’s Chief Information Officer.

A Statewide BI Application Advisory Council can play a key role in the successful deployment of business intelligence applications for the State. Specifically, the BI Advisory Council is seen as a critical component for maximizing the return on investment of BI applications used at the State.

Key roles of the Advisory Council include:

- Raising the awareness of the strategic importance of State data
- Ensuring appropriate compliance with privacy and security concerns
- Prioritizing BI applications based on statewide return on investment
- Facilitating change management best practices
- Ensuring that adequate funding has been obtained for BI deployment
- Ensuring that proper metrics are in place to measure the success of BI applications
- Ensuring that BI standards, policies, and procedures are followed consistently by State agencies.
- Ensuring that data quality standards are used and monitored.

- Assigning executive sponsors for each BI application.
- Ensuring that data stewards are assigned to be accountable for source data used for BI applications.

One of the most important functions of the Advisory Council will be to address cultural resistance to BI - particularly among first line directors and managers. Another key function is to make sure that Data Stewards are held accountable for the source data they have been entrusted to manage.

11. Utilize Consultants

CIBER recommends that the State utilize outside consultants from the BI and data integration/ETL software vendors and/or other BI consultancies as a means of reducing risks, speeding deployment, transferring knowledge, and reducing costs. Because of the State's limited BI and data integration/ETL expertise, use of outside consultants are recommended particularly for the initial implementation of the recommended BI, data warehousing, and data integration/ETL software and the establishment of the BI Competency Center.

The deployment of business intelligence and data integration software represents a complex endeavor that comes with associated risks with potentially damaging consequences. The use of experienced consultants is recommended to mitigate these risks by working with State staff.

Total outsourcing of the BI function is not recommended. It is critical to foster knowledge transfer and internal ownership of BI applications, consequently a proper mix of consultants to State staff is recommended to avoid over reliance on consultants.

12. Establish a Budget

Based on the recommendations in this report, CIBER recommends that the State establish a budget of approximately **\$1,240,000** to cover the recommendations offered in this report. This budget does not cover the cost of developing individual agency business intelligence applications. This budget also assumes that existing state personnel would be assigned to the BI Competency Center (four to five employees). This recommended budget is based on the following cost estimates:

- **BI Software– \$500,000.** Includes Enterprise licenses, administrator training, and software installation for both Cognos and SAS. This price includes Cognos licenses to support 5 named software administrators, 70 named report developers/power users and 1500 named BI consumers. It also includes a two processor SAS server license that does not have a limit on number of users (the SAS pricing is processor based with no charges for users). This cost also includes the purchase of the MITI Meta Integration® Model Bridge (MIMB) to integrate Meta data between SAS and Cognos.
- **Data Integration/ETL Software - \$240,000.** This would cover the one time purchase cost, first year support costs, administrator training, and installation costs for the purchase of data integration/ETL software (Informatica or SAS).

- **Data Warehouse Platform - \$270,000.** This would include the purchase of a data appliance (Netezza) that would include hardware and data storage with backup to support two terabytes of data warehouse data. This cost includes the one time purchase cost, first year maintenance fee (18%), installation, and onsite administrator training.
- **Training and Consulting - \$150,000.** This would cover the costs of using outside consultants as needed to establish the BI Competency Center and assist with issues and problems associated with building initial BI and ETL applications. This would also cover travel and registration costs for key BI Competency Center staff to attend two key BI conferences and two BI/ETL vendor user conferences. Employee training for the BI and ETL software is included in the cost of the BI and ETL software.
- **Server hardware - \$80,000.** This would cover the costs of two two-processor servers and related hardware to house the BI Software and the data integration/ETL software.
- **State employee costs (reassignment of four or five existing staff personnel).** It is recommended that four to five individuals be initially assigned to work part time in the BI Competency Center – one center director, one BI software administrator, one Data Integration/ETL administrator, one business analyst, and one database administrator (see Recommendation #7 above for additional details).

Additional details regarding the cost projections.

BI Software

The Cognos estimate breakdown is as follows:

Cognos 8 BI License Role	User counts	GSA pricing
BI Administrator	3	\$ 27,000
BI Professional (power user)	20	\$ 43,000
BI Business analyst	40	\$ 37,000
BI Consumer (2 CPU processor)	Unlimited consumers (limited by processor power)	\$263,000
	Total	\$370,000

Pricing includes first year support. The ETL administrator fee is \$52,500 and has been included in the ETL cost section (see below).

The SAS estimate breakdown is as follows:

SAS Enterprise BI Role	User Count	Estimated Pricing
2 CPU processor	Unlimited users (limited by processor power)	\$130,000

Pricing includes first year support

Data Integration Software: \$240,000

This cost estimate is based on the estimated software license fees and first year support for the purchase of the Informatica data integration software \$ and the Cognos ETL administrator license (\$52,500).

Server Hardware: \$80,000

This cost estimate is based on the purchase of two 2-4 processor servers one for BI software and one for the data integration software.

Data Warehouse Platform: \$270,000

This cost estimate is based on the purchase of a Netezza 5200 appliance configured to support 2 terabytes of data.

Appendix A: Current North Dakota BI Projects

The BI Study Team has developed an initial inventory of major State BI application projects. These projects include:

- **PeopleSoft Data Marts** - the State is licensed for PeopleSoft’s Enterprise Performance Management (EPM) software and five individual data marts (FIN, HRMS, CRM, Supply Chain, ESA) as well as a number of the “portals.” Included with PeopleSoft is licensing for Ascential ETL tool – which can only be used to import data for use by the Peoplesoft EPM software. The State has not deployed EPM or any of the data marts in a production environment. In conjunction with that effort, the implementation of PeopleSoft Enterprise Performance Management (EPM) was discussed. Revere Group was hired to develop a cost estimate for deploying EPM for consistent reporting of performance measures for five agencies – DOT, DHS, ITD, JSND and Protection and Advocacy. These agencies collect performance measures today from a variety of sources and use spreadsheets or Access databases to report the information.
- **PeopleSoft ad-hoc reporting** - the State replicates the PeopleSoft financial and HRMS operational database nightly to provide a query environment without impacting the operational database. PS Query is used by power users to generate reports. Data access and ad-hoc reporting are ongoing issues. PeopleSoft uses MS SQL as the backend database.
- **Game and Fish Management System** - Game and Fish developed a management reporting system with the help of SAS consultants. The system provides performance measures used by all levels of management in the agency and includes labor reporting, project information, and financial information. The SAS (www.sas.com) licensing purchased by Game and Fish is available for use by other agencies.
- **Disease Reporting** - the Dept of Health developed a system to summarize and report disease instances in the state. The system provides visual/graphical trend reporting using SAS. They have limited SAS licensing due to federal funding requirements.
- **Women, Infants and Children (WIC) Program the Dept of Health has implemented a WIC program that uses a vendor developed (CIBER) application that incorporates Microsoft Analysis Services. Analysis Services is a “free” component of Microsoft SQL Server.**
- **Bank of North Dakota (BND) – Core Banking** - BND is currently in the process of replacing their core banking system. They are implementing a COTS solution from ITI. The vendor has integrated COGNOS into the application to provide analytics.
- **Job Service North Dakota (JSND) – Labor Market Information** - JSND is currently in the process of procuring a replacement to its labor market information system (LMI). The LMI system provides a data warehouse and statistical employment reporting internally and via the web (www.jobsnd.com/data/index.html).
- **Workforce Safety and Insurance (WSI) – Fraud detection** WSI is interested in implementing a data warehouse for the purpose of tracking trends in injuries, analyzing rates, etc. They have Oracle tools for multi-dimensional analysis.
- **Department of Human Services (DHS) – Medstat** DHS out sources the Medstat application and transmits data to them monthly. The application is used to detect fraud in Medicare/Medicaid claims.

DHS is evaluating the need for a data warehouse and analytics as part of the replacement for their Medicaid Systems. The vendor they have selected, ACS, intends to use Cognos for development of the decision support system.

- **Tax Department** The Tax Department is in the process of implementing an integrated tax system call GenTax. Integrated within the product is a tool called “Discovery” that identifies potential matches to individuals in both internal and external files.
- **Courts – Data mart** The Judicial Branch uses a tool called Data Mirror to replicate their AS400 data to SQL Server to make the data more accessible to external users.
- **Department of Public Instruction (DPI) – Tetradata** DPI implemented a project to provide data warehouses for school districts across the state and one for DPI. Tetradata (www.tetradata.com) was the product implemented. Adoption has not been widespread and the project has been abandoned. Several larger districts are exploring alternative solutions. DPI has implemented a statewide student identification checking system to ensure student data is not duplicated in different districts.
- **ND University Systems (NDUS) – Round table performance measures** NDUS implemented a data warehouse (Oracle) several years ago to provide performance measure reporting. The warehouse has been abandoned due to lack of use.
- **FINDET** - FINDET is an interagency cooperative effort to provide follow-up information on North Dakota education and training programs. FINDET is also being used to meet the accountability measure requirements of the State legislature regarding training programs. FINDET uses a technique of data linking to match information on individuals in various educational and work programs to determine outcomes. Data linking uses existing data sources from State and Federal agencies, while maintaining the highest level of data security and requirements under Family Educational Rights and Privacy Act (FERPA) and applicable North Dakota laws and regulations.

Current BI Related State Technology Standards

Report Writer Tools:

- Business Objects Crystal Reports
- Jinfonet JReport
- Oracle Discoverer
- Oracle Reports

Enterprise Databases:

- Oracle
- MS SQL Server
- IBM DB2 UDB

Other Databases:

- MS Access

Appendix B: Business Intelligence Solution Requirements

The State BI Study Team developed the following high level Business Intelligence Solution Requirements in January 2006:

Prepared by: BI Study Team
Last Revision Date: January 19, 2006

Number	Type (Functional, Technical, Operational, Security)	High Level Requirement	Priority (Core, Essential, Desired)
1.	Technical	Off-load server/disk intensive query and reporting tasks to a separate infrastructure to avoid impacting the operational transaction processing system.	C
2.	Technical	Provide data models (e.g. star schema) and processes to optimize query and reporting functions.	C
3.	Functional	Provide an environment that allows for the speedy development of reports and queries with little technical training.	C
4.	Functional	Provide an “official” version of data that has been cleaned up and aligned to data reporting standards.	C
5.	Functional	Provide a single repository for data (structured and unstructured) from multiple transactional systems, running on any environment or database, so that a single report can be generated from multiple data sources.	C
6.	Security	Provide role based security to allow managers and others access to read-only data without giving them access to detailed transactional data.	C
7.	Functional	Capture historical (time-based) data to report trends in a manner that allows data transformation into consistent coding structures.	C
8.	Functional	Provide the capability to easily view data across multiple dimensions, i.e. time, program, agency, cost center, etc.	C
9.	Functional	Provide the capability to easily drill down from summary level into detailed data.	C
10.	Functional	Provide the ability to forecast data points based on statistical models.	C
11.	Functional	Provide the ability to do “what if” analysis based on modifications to assumptions.	C
12.	Security	Limit access to data based on roles. Security should be at the field / column / row level and based on the value of the data in the field, i.e. agency = 110.	C
13.	Functional	Provide the ability to view data via the web. Provide the ability to easily build web views based on user needs.	C
14.	Functional	Provide the ability to customize “report” views for individual users.	C
15.	Technical	Provide the ability to report the data using Crystal Reports.	C
16.	Functional	Provide the ability to enter data manually into the repository as well as from electronic interfaces.	C
17.	Functional	Provide the ability to schedule automatic uploads to the repository.	C
18.	Functional	Provide the ability to publish views directly to public web sites.	C

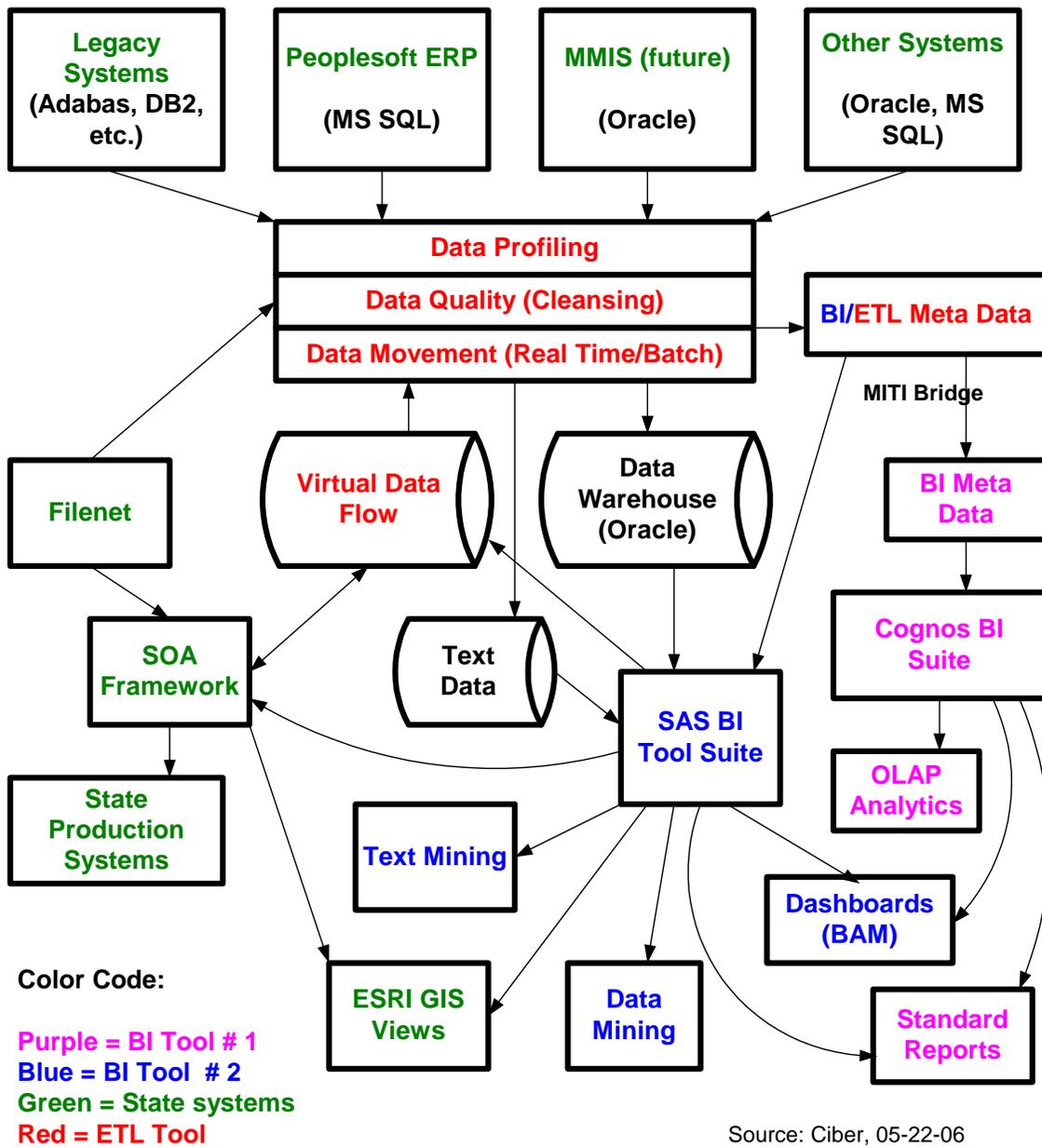
19	Functional	Provide the ability to correlate / match information on individuals (or other key data fields) in multiple databases and determine the probability of an exact match.	C
20	Capability	The state should have the core infrastructure in place to accommodate business intelligence solutions “on demand.”	C
21	Capability	The state should have staff trained in the development and deployment of business intelligence solutions from a functional and technical perspective.	E

Definition of 'Priority'	
Term	Definition
Core	Requirements without which the product may as well not be developed at all; it will be of no use to most Customers without these.
Essential	Requirements for which a short-term work-around could be developed (or for which an old process can be used for a little while longer), but over the long run, the requirements have to be there.
Desired	Requirements that are the "bells and whistles" which may be precious to certain constituencies, but without which the product will function just fine.

Appendix C: North Dakota BI Enterprise Architecture Chart (a sample view of potential enterprise application integration)

State of North Dakota

Business Intelligence Architecture - Dual BI Scenario



Appendix D: Glossary

From www.DataWarehousing.com:

A

Agent	An application that searches data and sends an alert when a certain situation occurs. (See ALERT)
Aggregate data	Individual data items, data groups, arrays, tables etc. that can be assembled to form a whole.
Alerts and Alarms	Messages sent automatically by a computer system when an AGENT identifies a certain situation. For example, if stock of an item in a warehouse drops to a certain level key personnel can be immediately informed.
Algorithm	A set of statements organized to solve a problem in a finite number of steps
Analytical Processing	The usage of the computer to produce an analysis for management decision, usually involving trend analysis, drill down analysis, demographic analysis, profiling, etc.
Architecture Phase	The establishment of the framework, scope and standards and procedures for a data warehouse at the enterprise level.
Atomic level data	Data with the lowest level of granularity. Atomic level data sits in a data warehouse and is time variant (i.e., accurate as of some moment in time now passed).
Attribute	A property or characteristic of an application entity. For example, the attributes of an EMPLOYEE entity in a business application may be: IDFirstname Lastname Job_Title Email_ID An attribute usually represents a column in a table in a relational database, or a field in a file.
Audit Trail	Recording of any changes made to specific data. Details can include date and time of change, how the change was detected, reason for the change and before-and-after data values.

B

Binary Search	A dichotomizing search with steps in which the sets of remaining items are partitioned into two equal parts.
Bit Map	A specialized form of an index indicating the existence or non-existence of a condition for a group of blocks or records.
Bus	The hardware connection that allows data to flow from one component to another.
Business Intelligence Tools	Software that allows business users to see and use large amounts of complex data.

C

Canonical model	A data model that represents the inherent structure of data without regard to either individual use or hardware or software implementation.
Cardinality	no of unique rows divided by total no of columns
Cell	A single point in a CUBE.
Conceptual Schema	A consistent collection of data structures expressing the data needs of the organization. This schema is a comprehensive, base level, and logical description of the environment in which an organization exists, free of physical structure and application system considerations.
Condensation	The process of reducing the volume of data managed without reducing the logical consistency of the data.
Connector	A symbol used to indicate that one occurrence of data has a relationship with another occurrence of data. Connectors are used in conceptual data base design and can be implemented hierarchically, relationally, in an inverted fashion, or by a network.
Contention	The condition that occurs when two or more programs try to access the same data at the same time.
Cooperative Processing	The ability to distribute resources (programs, files and data bases) across the network.
Corporate Data	All the databases of the company. This includes legacy systems, old and new transaction systems, general business systems, client/server databases, data warehouses and data marts.
Corporate Information Warehouse (CIF)	The architectural framework that houses the ODS, data warehouse, data marts, i/t interface, and the operational environment. The CIF is held together logically by metadata and physically by a network such as the Internet.
Cube – (also Hypercube, Multi-dimensional Cube)	The fundamental structure for information in an OLAP system. A structure that stores multi-dimensional information, having one CELL for each possible combination of dimensions.

D

Data	Facts, concepts, or instructions that a computer records, stores and processes. Used in conjunction with INFORMATION SYSTEMS, “raw data” is organized in such a way that people can understand the results.
Data Cleansing	Removing errors and inconsistencies from data being imported to a data warehouse.
Data Dictionary	A software tool for recording the definition of data, the relationship of one category of data to another, the attributes and keys of groups of data, and so forth.

Data Driven Development	The approach to development that centers around identifying the commonality of data through a data model and building programs that have a broader scope than the immediate application.
Data Driven Process	A process whose resource consumption depends on the data on which it operates.
Data Mart	A department-specific data warehouse. A) Independent – fed from legacy systems within the department B) Dependent – fed from the enterprise data warehouse (preferred)
Data Mining	The process of finding hidden patterns and relationships in data. For instance, a consumer goods company may track 200 variables about each consumer. There are scores of possible relationships among the 200 variables. Data mining tools will identify the significant relationships.
Data Scrubbing	Removing errors and inconsistencies from data being imported into a data warehouse.
Data Transformation	The modification or alteration of data as it is being moved into the data warehouse.
Data Type	A data type defines the type of data stored in a specific database column, such as date, numeric or character data. Significant differences in data types exist between different platforms' databases.
Data Warehouse	A data warehouse is a subject oriented, integrated, non volatile, time variant collection of data. The data warehouse contains atomic level data and summarized data specifically structured for querying and reporting.
Data Warehousing	An enterprise-wide implementation that replicates data from the same publication table on different servers/platforms to a single subscription table. This implementation effectively consolidates data from multiple sources.
Database Schema	The logical and physical definition of a database structure.
Date/Time Stamp	A stamp added by an application that identifies a task or activity by the date and time it was initiated and/or completed. This can appear as part of a transaction log, message queue content in job logs.
Decentralized Database	A centralized database that has been partitioned according to a business or end-user defined subject area. Typically ownership is also moved to the owners of the subject area.
Decentralized Warehouse	A remote data source that users can query/access via a central gateway that provides a logical view of corporate data in terms that users can understand. The gateway parses and distributes queries in real time to remote data sources and returns result sets back to users.
Decision Support Systems (DSS)	Software that supports exception reporting, stop light reporting, standard repository, data analysis and rule-based analysis. A database created for end-user ad-hoc query processing.
Denormalization	The technique of placing normalized data in a physical location that optimizes the performance of the system.
Derived Data	Data whose values are determined by equations or algorithms.
Dimension	A Dimension is typically a qualifiable and text value, such as a region, product line, and includes date values. It defines the secondary headings or labels that make up the body of the report. Each of the dimensions is repeated within each group. Usually, you use items containing text values (for example, Year or item type) for table dimensions. For example, if you

select Item Type to be your table dimension, Item Type is a dimension within each group header. Under the dimension "Item Type," appears the name of each kind of item (for example, CD ROM, or HARD Drive) and corresponds to the fact. A fact is a quantifiable value, such amount of sales, budget or revenue.

Drill Down/Up The ability to move between levels of the hierarchy when viewing data with multiple levels.
 A) Drill down – changing a view to a greater level of detail
 B) Drill up – changing a view to a greater level of aggregation.

E

EDI (Electronic Data Interchange) A standard format for exchanging business data.

Encryption The transformation of data from a recognizable format to a form unrecognizable without the algorithm used for the transformation.

ETL (Extract, Transform and Load) ETL refers to the process of getting data out of one data store (extract), modifying it (transfer), and inserting it into a different data store (load).

ETT ETL is sometimes referred as ETT- Extraction, Transformation and Transportation. It is a series of batch interface between the systems.

Executive/Enterprise Information Systems (EIS) Tools programmed to provide canned reports or briefing books to top-level executives. They offer strong reporting and drill-down capabilities. Today these tools allow ad-hoc querying against a multi-dimensional database, and most offer analytical applications along functional lines such as sales or financial analysis. (Also known as Decision Support System.)

Extendibility The ability to easily add new functionality to existing services without major software rewrites or without redefining the basic architecture.

External Schema A logical description of a user's method of organizing and structuring data.

F

Fact Table The tables which are extracted from heterogeneous sources and used in the Data Warehouse

Factless Fact A fact table without any metrics in it

Flat File A collection of records containing no data aggregates, nested repeated data items, or groups of data items.

Functional Decomposition The division of operations into hierarchical functions that form the basis for procedures.

G

Global Business Models	Provides access to information scattered throughout an enterprise under the control of different divisions or departments with different databases and data models. This type of data warehouse is difficult to build because it requires users from different divisions to come together to define a common data model for the warehouse.
Granularity	The level of detail of the data stored in a data warehouse.

H

Heterogeneous Environment	Within an enterprise, a network of different types of servers and databases.
Heuristic	The mode of analysis in which the next step is determined by the results of the current step of analysis.
Hierarchy	The organization of data into a logical tree structure.
Homogeneous Environment	Within an enterprise, a network consisting of the same type of servers and databases.
Horizontal Distribution	The splitting of a table across different sites by rows. With horizontal distribution rows of a single table residing at different sites in a distributed data base network.
Hub and Spoke Configuration	A configuration of interconnected systems where a single system (the hub) acts as the central point for exchanging data with and between the other systems (spokes).
Huffman Code	A code for data compaction in which frequently used characters are encoded with fewer bits than infrequently used characters.
HyperCube	See CUBE.

I

Indexing	Fastest searching records
Information	Data that has been processed in such a way that it can increase the knowledge of the person who receives it.
Information Systems Architecture	The authoritative definition of the business rules, systems structure, technical framework, and product backbone for business information systems.
Instance	A set of values representing a specific entity belonging to a particular entity type.
Integrity	A set of values representing a specific property of a data base that ensures that the data contained in the data base is accurate and consistent as possible.

Intelligent Data Base	A data base that contains shared logic as well as shared data and automatically invokes that logic when the data base is accessed. Logic, constraints, and controls relating to the use of data are represented in an intelligent data model.
Interleaved Data	Data from different tables mixed into a simple table space where is commonality of physical co-location based on a common key value.
Iterative Analysis	The mode of processing in which the next step of processing depends on the results obtained by the existing step in execution.

J

Join	An operation that takes two relations as operands and produces a new relation by concatenating the tuples and matching the corresponding columns when a stated condition holds between the two.
------	---

K

Key Compression	A technique for reducing the number of bits in keys; used in making indexes occupy less space.
-----------------	--

L

Latency	Often used to mean any delay or waiting that increases real or perceived response time beyond the response time desired.
Load	After extracting, cleaning and transforming, data must be loaded into the warehouse. Additional preprocessing may still be required: checking integrity constraints; sorting; summarization, aggregation and other computation to build the derived tables stored in the warehouse; building indices and other access paths; and partitioning to multiple target storage areas. Load utilities can be used for these operations.
Lockup	The event that occurs when update is done against a data base record and the transaction has not yet reached a commit point.
Logging	The automatic recording of data with regard to the access of the data, the updates to the data, etc.
Logical Representation	A data view or description that does not depend on a physical storage device or a computer program.

M

Main Storage Data Base (msdb)	A data base that resides entirely in main storage. Such data bases are very fast to access, but require special handling at the time of update. MSDB's can only manage a small amounts of data.
Maximum Transaction Arrival Rate (MTAR)	The rate of arrival of transactions at the moment of peak period processing.
MDDB	Multi Dimensional DataBase
Metadata or Meta Data	Metadata is data about data. Examples of metadata include data element descriptions, data type descriptions, attribute/property descriptions, range/domain descriptions, and process/method descriptions. The repository environment encompasses all corporate metadata resources: database catalogs, data dictionaries, and navigation services. Metadata includes things like the name, length, valid values, and description of a data element. Metadata is stored in a data dictionary and repository. It insulates the data warehouse from changes in the schema of operational systems.
Metadata Synchronization	The process of consolidating, relating and synchronizing data elements with the same or similar meaning from different systems. Metadata synchronization joins these differing elements together in the data warehouse to allow for easier access.
Metalanguage	A language used to specify other languages.
Methodology	A system of principles, practices, and procedures applied to a specific branch of knowledge.
Mid-Tier Data Warehouses	To be scalable, any particular implementation of the data access environment may incorporate several intermediate distribution tiers in the data warehouse network. These intermediate tiers act as source data warehouses for geographically isolated sharable data that is needed across several business functions.
Middleware	A communications layer that allows applications to interact across hardware and network environments.
Migration	The process by which frequently used items of data are moved to more readily accessible areas of storage and infrequently used items of data are moved to less readily accessible areas of storage.
MOLAP	OLAP on Multidimensional models. In MOLAP servers, Data warehouses directly store multidimensional data in special data structures (eg. arrays) and implement the OLAP operations over these special data structures.
Multilist Organization	A chained file organization in which the chains are divided into fragments and each fragment is indexed. This organization of data permits faster access to the data.

N

Natural Join	A join in which the redundant logic components generated by the join are removed.
Network Model	A data model that provides data relationships on the basis of records or groups of records (ie. sets) in which one record is designated as the set owner, and a single member record can belong to one or more sets.

Nonprocedural Language	Syntax that directs the computer as to what to do, not how to do it. Typical nonprocedural languages include RAMIS, FOCUS, NOMAD, and SQL.
Normalization	Normalization is a step-by-step process of removing redundancies and dependencies of attributes in a data structure. The condition of the data at completion of each step is described as a "normal form." Thus, when normalizing we talk about data as being in the first normal form, the second normal form, etc. Normalization theory identifies normal forms up to at least the fifth normal form, plus an adjunct form known as Boyce-Codd Normal Form (BCNF). The first three forms are sufficient to meet the needs of warehousing data models.

O

OLAP (On-Line Analytical Processing)	Describes the systems used not for application delivery, but for analyzing the business, e.g., sales forecasting, market trends analysis, etc. These systems are also more conducive to heuristic reporting and often involves multidimensional data analysis capabilities.
OLTP (OnLine Transaction Processing)	Describes the activities and systems associated with a company's day-to-day operational processing and data (order entry, invoicing, general ledger, etc.).
Operational Data Store (ODS)	The form that data warehouse takes in the operational environment. Operational data stores can be updated, do provide rapid and consistent time, and contain only a limited amount of historical data.
Overflow	The condition in which a record or a segment cannot be stored in its home because the address is already occupied.

P

Parallel Data Organization	An arrangement of data in which the data is spread over independent storage devices and is managed independently.
Parallel Search Storage	A storage device in which one or more parts of all storage locations are queried simultaneously for a certain condition or under certain parameters.
Parsing	The algorithm that translates syntax into meaningful machine instructions. Parsing determines the meaning of statements issued in the data manipulation language.
Partition	A segmentation technique in which data is divided into physically different units. Partitioning can be done at the application or the system level.
Performance	The length of time from the moment a request is issued until the first of the results of the request are received.
Periodic Discrete Data	A measurement or description of data taken at a regular time interval.
Prefix Data	Data in a segment or a record used exclusively for system control, usually unavailable to the user.

Primitive Data	Data whose existence depends on only a single occurrence of a major subject area of the enterprise.
Privilege Descriptor	A persistent object used by a DBMS to enforce constraints on operations.
Projection	An operation that takes one relation as an operand and returns a second relation that consists of only the selected attributes or columns, with duplicate rows eliminated.
Proposition	A statement about entities that asserts or denies that some condition holds for those entities.

Q

Query Language	A language that enables an end user to interact directly with a DBMS to retrieve and possibly modify data managed under the DBMS.
----------------	---

R

Record	An aggregation of values of data organized by their relation to a common key.
Recovery	The restoration of the database to an original position or condition, often after major damage to the physical medium.
Redundancy	The practice of storing more than one occurrence of data.
Referential Integrity	The facility of a DBMS to ensure the validity of a predefined relationship.
Refresh	Refreshing a warehouse consists in propagating updates on source data to correspondingly update the base data and derived data stored in the warehouse. Two sets of issues to consider; when to refresh and how to refresh. Refresh policy is set by warehouse administrator, depends on user needs and traffic and may be different for different sources.
Replication	The physical copying of data from one database to another.
Reporting	The process of translating data to presentation formats via a pre-defined or ad-hoc queries.
ROLAP	Relational OLAP. Data warehouses that are implemented on standard or extended relational DBMSs, called Relational OLAP (ROLAP) servers. These servers assume that data is stored in relational databases.
Roll up	To increase or acquire by successive accumulations
Rolling Summary	A form of storing archival data where the most recent data has the lowest level of details stored and the older data has higher levels of details stored.

S

Schema	The logical organization of data in a database.
Scope of Integration	The formal definition of the boundaries of the system being modeled.
Sequential File	A file in which records are ordered according to the values of one or more key fields.
Serial File	A sequential file in which records are physically adjacent, in sequential order.
Slowly Changing Dimensions	The approaches involving maintaining a list or history by adding related rows or new columns, or simply ignoring the problem by retaining the only the current data. Type I, Type II, Type III
Snowflake Schema	A snowflake schema is a set of tables comprised of a single, central fact table surrounded by normalized dimension hierarchies. Each dimension level is represented in a table. Snowflake schema implement dimensional data structures with fully normalized dimensions. Star schema are an alternative to snowflake schema.
Star Schema	A star schema is a set of tables comprised of a single, central fact table surrounded by de-normalized dimensions. Each dimension is represented in a single table. Star schema implement dimensional data structures with de-normalized dimensions. Snowflake schema are an alternative to star schema.
Surrogate Key	It has system-generated artificial primary key values, which allows the maintenance of historical records in the Data Warehouse more effectively.

T

Table	A relation that consists of a set of columns with a heading and a set of rows.
Time Variant Data	Data whose accuracy is relevant to some one moment in time.
Top down methodology	Involves in building a data warehouse first and then building data marts.
Transaction Processing	The activity of executing many short, fast running programs, providing the end user with consistent two or three second response time.
Transition Data	Data possessing both primitive and derived characteristics; usually very sensitive to the running of the business.

Endnotes

¹ Based on a survey of 1,400 CIO's reported by the Gartner Group in 2006

² Feb/March 2005 Survey conducted by Better Management Inc.

³ Feb/March 2005 Survey conducted by Better Management Inc.

⁴ Blumberg, R. and Atre, S. "The problem with unstructured data." *DM Rev* February 2003

⁵ From SAS press release, August 10, 2005

⁶ From Cognos Product Informational Material 2006